

THESIS / THÈSE

MASTER EN BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE ET CELLULAIRE

Sélection assistée par marqueurs chez *Cichorium intybus* L. : recherche de marqueurs moléculaires ciblant les gènes de floraison

Dognaux, Sophie

Award date:
2009

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX
NAMUR**

Faculté des Sciences

**Sélection assistée par marqueurs chez *Cichorium intybus* L. :
recherche de marqueurs moléculaires ciblant les gènes de floraison**

**Mémoire présenté pour l'obtention
du grade académique de master en Biochimie et Biologie moléculaire et cellulaire**

Sophie Dognaux

Janvier 2009

**Sélection assistée par marqueurs chez *Cichorium intybus* L.:
recherche de marqueurs moléculaires ciblant les gènes de floraison**

DOGNAUX Sophie

Résumé

La chicorée industrielle (*Cichorium intybus* L.) est une plante bisannuelle qui nécessite une exposition au froid et à de longues photopériodes pour atteindre sa phase reproductive. Cependant, dans certaines conditions, la plante monte et fleurit dès la première année, réduisant ainsi les ressources carbonées exploitables qui sont stockées dans ses racines. L'objectif de ce mémoire était de contribuer au développement de marqueurs moléculaires permettant d'améliorer et d'accélérer la sélection de plantes résistantes à la montaison.

Pour débiter ce travail, vingt-et-un gènes candidats ont été sélectionnés sur base de leur fonction dans les différentes voies de floraison décrites chez *Arabidopsis thaliana*. L'amplification de ces gènes candidats a été effectuée sur une core collection d'ADNg issus des vingt-neuf variétés à la base de la sélection de *Cichorium intybus* L. La recherche du polymorphisme génétique de type SNP (Single Nucleotide Polymorphism) au sein de quatorze gènes candidats a alors été réalisée. La détection des SNP a été effectuée, dans un premier temps, par des analyses EcoTILLING sur des pools d'échantillons puis, finalement, sur des échantillons individuels. Les SNP les plus fréquemment détectés ont été génotypés par séquençage, en se basant sur l'hypothèse qu'une fréquence ancestrale importante augmente la probabilité du SNP d'être toujours présent à l'heure actuelle.

L'énorme diversité génétique observée au sein de la core collection représente ainsi une analyse préliminaire intéressante. En effet, l'objectif final est la mise en évidence de SNP qui auraient été conservés et qui pourraient donc être utilisés dans la sélection d'individus issus de variétés actuelles. Pour accélérer cette sélection, une technique d'analyse à haut débit dérivée de la méthode du Golden Gate Assay devra encore être mise au point. Celle-ci pourra alors être utilisée pour des études ultérieures en génomique structurale (cartographie génétique) et fonctionnelle (études d'association) afin de concevoir des marqueurs moléculaires exploitables en sélection.

Mémoire de master en Biochimie et Biologie moléculaire et cellulaire
Janvier 2009

Promoteur: P. Van Cutsem **Co-promoteur:** N. Dauchot

Je voudrais tout d'abord remercier Mr Messiaen de m'avoir accueillie au sein de son laboratoire, où bonne humeur et humour noir riment avec rigueur et travail.

Je voudrais également remercier mon promoteur, Mr Van Cutsem, de m'avoir enseigné que la ténacité et la droiture sont des valeurs sûres dans la vie.

J'aimerais ensuite remercier Nicolas de m'avoir initiée au travail de laboratoire et de m'avoir transmis une méthode de travail rigoureuse, fruit de ses nombreuses années d'expérience. Merci également pour tous ces conseils avisés et discussions critiques partagés au cours de cette année.

Un grand merci également à tous les membres de l'URBV de m'avoir si bien accueillie et intégrée dans la vie du labo. Merci aussi à Céline, Monique et Adrien d'avoir apporté encore d'autres horizons au sein du labo.

Et puis un merci tout particulier à mes trois compères mémorants pour tous ces bons (et moins bons) moments passés ensemble et tout ce que chacun d'entre vous à pu m'apprendre. Merci à toi Ann-Florence de m'avoir transmis quelques (brèves) notions d'Allemand et de m'avoir enfin fait déguster ce fameux *Schnaps*. Merci à toi Stéphanie de m'avoir appris qu'une journée pouvait compter plus de 24 heures quand on est vraiment passionné par ce que l'on fait. Enfin, merci à toi Boris de m'avoir partagé ta passion (viscérale) pour les SNP et la science en générale. Cette année n'aurait pas été pareille sans vous...

Merci également à toute la cuvée BIO 2008-2009 pour ces cinq années de folie passées ensemble en tant que « cobayes » ! On ne se quitte pas encore mais je voudrais déjà vous souhaiter bonne route pour la fin de cette année et bon vent pour les années à venir...

J'aimerais finalement remercier tous mes amis et toute ma famille d'avoir toujours été présents et d'avoir rendu, chacun à votre façon, mes instants en dehors de l'unif tous aussi uniques et délirants. Je vous assure que la fin de ce mémoire marque le début d'une époque encore plus folle à partager.

Enfin, je voudrais remercier d'avance les membres de mon jury pour le temps qu'ils consacreront à la lecture de ce manuscrit.

Sophie

Liste d'abréviations

ADN	Acide DésoxyriboNucléique
ADNc	ADN complémentaire
ADNg	ADN génomique
AFLP	Amplification Fragment Length Polymorphism
AG	AGAMOUS
AP	APETALA
APRR	Arabidopsis PSEUDO-RESPONSE REGULATOR
ARN	Acide RiboNucléique
ARNm	ARN messenger
ASO	Allele Specific Oligonucleotide
ATP	Adénosine TriPhosphate
BSA	Bovine Serum Albumin
CCA1	CIRCADIAN CLOCK-ASSOCIATED 1
CDPK	Calcium-Dependant Protein Kinase
CIAP	Calf Intestinal Alkaline Phosphatase
CO	CONSTANS
COP1	Constitutive Photomorphogenic 1
CpG	Cytosine-phosphate-Guanine
CRY	CRYPTOCHROME
D.O.	Densité Optique
DGGE	Denaturing Gradient Gel Electrophoresis
DHPLC	Denaturing High Performance Liquid Chromatography
DNTP	DiNucléotide TriPhosphate
EDTA	Ethylène Diamine Tétra-Acétate
ELF	EARLY-FLOWERING
Endo-1	Endonucléase-1
EST	Expressed Sequence Tag
FAD	Flavine Adénine Dinucléotide
FES1	FRIGIDA-ESSENTIAL 1
FKF1	FLAVIN-BINDING KELCH REPEAT F-BOX 1
FLC	FLOWERING LOCUS C
FLD	FLOWERING LOCUS D
FLK	FLOWERING LOCUS K
FMN	Flavine MonoNucléotide
FRI	FRIGIDA
FT	FLOWERING LOCUS T
GA	Gibberellic Acid
GAI	GIBBERELLIC ACID INSENSITIVE
GI	GIGANTEA
HDAC	Histone Désacétylase
HOS1	High expression of Osmotic Stress-regulated gene expression 1
HP1	Heterochromatin Protein 1

IPTG	Isopropyl-Beta-Thio-Galactoside
ISSR	Inter-Simple Sequence Repeat
LD	LUMINIDEPENDENS
LFY	LEAFY
LHY	LATE-ELONGATED HYPOCOTYL
LOV	Light-Oxygen-Voltage
LSO	Locus Specific Oligonucleotide
MCS	Multiple Clonage Site
NLS	Nuclear Localisation Sequence
pb	paire de base
PCI	Phénol : Chloroforme : alcool Isoamylique
PCR	Polymerase Chain Reaction
PEPB	Phosphatidyl-Ethanolamines Binding Protein
PFT1	PHYTOCHROME AND FLOWERING TIME 1
PHD	Plant Homeo-Domain
PHR	Photolyase-Related
PHY	PHYTOCHROME
PI	PISTILLATA
PIF3	Phytochrome Interacting Factor 3
QTL	Quantitative Trait Loci
REF6	RELATIVE OF EARLY-FLOWERING 6
RFLP	Restriction Fragment Length Polymorphism
RGA	REPRESSOR OF GA1-3
RGL1	RGA-LIKE 1
Rnase	Ribonucléase
SAP	Shrimp Alkaline Phosphatase
SEP	SEPATALA
SL	Smart Ladder
SLS	Sample Loading Solution
SNP	Single Nucleotide Polymorphism
SOC1	SUPPRESSOR OF OVEREXPRES-SION OF CO1
spp	species
SSCP	Single-Strand Conformation Polymorphism
SUF4	SUPPRESSOR OF FRIGIDA-4
TAIR	The Arabidopsis Information Resource
TILLING	Targeting Induced Local Lesions in Genomes
TOC1	TIMING OF CAB1
VIL1	VIN3-like1
VIN3	VERNALIZATION-INSENSITIVE 3
VIP4	VERNALIZATION INDEPENDENCE 4
VRN	VERNALIZATION
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
ZTL	ZEITLUPE

Table des matières

Introduction

1. Chicorée et montaison.....	1
2. Gène candidat.....	2
3. Gènes candidats dans les voies de floraison	2
3.1. Voie dépendante de la photopériode	3
3.2. Voie dépendante de la vernalisation.....	7
3.3. Voie autonome.....	10
3.4. Voie dépendante de l'acide gibbérellique	11
3.5. Intégration des différentes voies de floraison.....	12
4. Polymorphisme des gènes candidats.....	14
4.1. Polymorphisme et marqueurs moléculaires.....	14
4.2. Recherche de SNP	15
4.3. Génotypage de SNP.....	16
4.4. Utilisation des SNP en génomique	17
5. Objectifs du mémoire.....	18

Matériels et méthodes

1. Extraction de l'ADNg des feuilles de chicorée.....	19
2. Design d'amorces et amplification des gènes candidats.....	20
3. Clonage des gènes candidats.....	20
3.1. Préparation de cellules thermocompétentes	20
3.2. Transformation bactérienne par choc thermique	21
3.3. Maxiprep.....	22
3.4. Linéarisation du vecteur	23
4. Purification des produits PCR et du vecteur	23
5. Etape de restriction-ligation.....	24
6. Transformation bactérienne par choc thermique	24
7. Screening.....	24
8. EcoTILLING.....	25
9. Miniprep.....	26
10. Séquençage	27

Résultats et discussion

1. Design du pool d'ADNg	28
2. Sélection des gènes candidats	28
3. Conception d'amorces	28
4. Clonage des gènes candidats.....	29
5. Analyses EcoTILLING	29
5.1. Analyses EcoTILLING en pools.....	30
5.2. Analyses EcoTILLING en « un contre un »	33
6. Analyses de séquençage.....	36
6.1. Première analyse.....	36
6.2. Deuxième analyse.....	40

Conclusion et perspectives.....	46
--	-----------

Bibliographie.....	49
---------------------------	-----------

1. Chicorée et montaison

La chicorée industrielle (*Cichorium intybus* L.) appartient à la famille des *Asteraceae* ou *Compositae*. La dénomination *Astéracées* vient du genre *Aster* (marguerite), tandis que la dénomination *Composées* caractérise le type d'inflorescence de cette famille. En effet, les dicotylédones appartenant à cette famille présentent une inflorescence particulière au sein de laquelle plusieurs fleurs sont regroupées pour former un capitule similaire à une seule grosse fleur. Ce type d'inflorescence (*pseudanthia*) n'est présent que chez quelques autres familles d'angiospermes. La famille des *Astéracées* comprend principalement des plantes herbacées (chardon, pissenlit, marguerite, pâquerette, edelweiss, tournesol), mais on peut également y trouver des arbres ou des arbustes. D'autres sont cultivées comme plantes ornementales (chrysanthème, gerbera, dahlia) ou pour la consommation de leurs racines (salsifis, topinambour) ou de leurs feuilles (laitue, endive, scarole). Les *Astéracées* sont des plantes plutôt cosmopolites mais préfèrent les régions tempérées ou les montagnes tropicales¹. Le genre *Cichorium* compte six variétés dispersées, à l'état sauvage, dans de nombreuses régions d'Europe et d'Asie (Bais et al., 2001).

Autrefois, la chicorée industrielle était utilisée pour la production de papyrus. Par la suite, la chicorée fut plutôt utilisée pour ses vertus thérapeutiques, comme de nombreux autres genres d'*Astéracées* encore utilisées en phytothérapie. La chicorée présente également de nombreuses propriétés nutritives, d'où la consommation de ses feuilles mangées crues en salade ou de ses racines qui, une fois hachées et torréfiées, sont utilisées comme substitut de café. C'est à la fin du 20^{ème} siècle que la culture de cette plante connut un essor important, suite à son utilisation pour la production d'inuline. En effet, les *Astéracées* contiennent de nombreux métabolites distribués dans toute la plante, dont certains principalement dans la racine. Les racines de chicorée peuvent ainsi contenir jusqu'à 20 % d'inuline, un polymère de β -2,1-fructose constituant une réserve d'énergie pour les plantes (**Figure 1**). La majorité de la matière organique assimilée au cours de la photosynthèse est utilisée pour la synthèse de saccharose qui est ensuite transporté dans toute la plante. Dans la racine, ce saccharose peut être transformé et stocké sous forme d'oligosaccharides tels que l'inuline. Celle-ci peut alors être utilisée pour permettre aux plantes de survivre en cas de manque d'eau, suite au froid ou à la sécheresse. En effet, l'inuline peut être convertie en fructose et glucose par hydrolyse chimique ou enzymatique (Cornelis, 2003). Dans l'industrie agro-alimentaire, l'inuline peut être utilisée en tant que fibre alimentaire ou édulcorant à faible niveau énergétique, ainsi qu'en tant que composant à effet prébiotique.

La chicorée industrielle a une taille comprise entre 80 et 90 cm de haut et présente une racine charnue pouvant atteindre 75 cm. Cette plante vivace et touffue forme de petites fleurs de couleur bleu-lavande (Bais et al., 2001 – **Figure 2**). La chicorée cultivée pour ses racines préférera un climat chaud et humide caractéristique des régions méridionales, tandis que celle cultivée pour la production de graines préférera un climat tempéré et sec (Cornelis, 2003). La chicorée est cependant une plante robuste pouvant également résister au gel. Ainsi, cette plante bisannuelle nécessite une exposition au froid (vernalisation) et à de longues photopériodes pour atteindre sa phase reproductive. Pour l'initiation de sa floraison, la plante (ou ses graines) doit être exposée à de basses températures ($\sim 4^{\circ}\text{C}$) pendant plusieurs jours (~ 3 semaines). Ce processus est appelé vernalisation. Ainsi, en conditions naturelles, la chicorée produit une rosette de feuilles vertes au cours de la première année puis, suite à la vernalisation subie pendant l'hiver, elle fleurit au cours de l'été de sa deuxième année.

¹ Cette section est inspirée de la banque de données de la « Texas A&M University » (<http://www.csdl.tamu.edu/FLORA/tamuherb.htm>)

Cependant, dans certaines conditions, la chicorée produit une hampe florale dès la première année. Ce phénomène, appelé montaison, est caractérisé par une augmentation du nombre de cellules et par l'élongation des tiges qui porteront, par la suite, les fleurs et les graines (Demeulemeester et al., 1999). Lors de ce processus, l'inuline accumulée dans la racine est partiellement hydrolysée afin de fournir des ressources carbonées à la plante en développement. De ce fait, la valeur énergétique, et donc économique, de ces racines est réduite, ce qui n'est pas d'un grand intérêt pour les cultivateurs. Le but de ce travail est d'améliorer la sélection des plantes résistantes à la montaison, via l'approche gène candidat et la sélection assistée par marqueurs.

2. Gène candidat²

La plupart des caractères d'intérêt agricole sont des QTL (Quantitative Trait Loci) définis comme des caractères quantitatifs dont l'hérédité est polygénique. Ainsi, un phénotype particulier peut dépendre de plusieurs gènes et varier continuellement suite à l'interaction de ces différents gènes entre eux et avec l'environnement. Les QTL ne correspondent pas toujours aux gènes eux-mêmes mais peuvent y être étroitement associés. A cause de cette multiplicité de gènes définissant un caractère complexe, leur impact potentiel sur la variation phénotypique et leur localisation sur les cartes génétiques restent imprécis. C'est pourquoi, l'approche gène candidat semble la solution la plus adaptée pour la caractérisation de ces QTL.

En génétique végétale, un gène candidat est défini comme un gène suspecté d'être impliqué dans l'expression d'un trait particulier. Deux types de gènes candidats peuvent être distingués : les gènes candidats positionnels et les gènes candidats fonctionnels. Les gènes candidats positionnels sont des gènes localisés sur des cartes de liaison et qui sont étroitement liés à des loci contrôlant un phénotype donné. Le choix de ces gènes candidats est basé sur leur co-ségrégation avec les loci affectant le phénotype. Les gènes candidats fonctionnels sont des gènes suspectés d'affecter un phénotype suite à leur intervention dans les voies biochimiques et/ou physiologiques liées à celui-ci. Donc, le choix de ces gènes candidats est basé sur leur rôle potentiel dans la variation phénotypique. En effet, les mutations localisées dans la région codante, non codante ou promotrice d'un gène pourraient être responsables de la variabilité phénotypique d'un trait donné. C'est pourquoi on peut définir un gène candidat, d'un point de vue plus général, comme un gène présentant du polymorphisme moléculaire soit génétiquement lié à des loci connus, soit statistiquement associé aux variations phénotypiques du trait étudié.

3. Gènes candidats dans les voies de floraison

Pour l'amélioration de la résistance à la montaison, les gènes candidats fonctionnels impliqués dans les différentes voies de floraison semblent être des cibles intéressantes. Étant donné que la floraison a été bien étudiée chez la plante modèle *Arabidopsis thaliana*, la sélection des gènes candidats chez la chicorée a été réalisée sur base des informations disponibles à propos des quatre voies principales de floraison connues chez *A. thaliana*.

² Cette section est inspirée de l'article de Pflieger et al., 2001

3.1. Voie dépendante de la photopériode

Les plantes acquièrent l'énergie indispensable à leur survie uniquement par la lumière (Bae et al., 2008). C'est pourquoi la photopériode est l'un des facteurs environnementaux les plus importants dans le contrôle du développement végétal. La lumière permet aux plantes de s'adapter à divers environnements et, dans nos régions tempérées, elle représente un facteur clé pour l'initiation de la floraison. Il est en effet essentiel que les plantes coordonnent leur floraison et leur reproduction sexuelle afin que ces deux processus se produisent au moment le plus approprié de l'année (Mouradov et al., 2002). Chez la plupart des organismes, la lumière est perçue par de multiples photorécepteurs dont le rôle est de transmettre le signal aux voies de régulation en aval (Bae et al., 2008). Chez les végétaux, l'ensemble du spectre d'absorption perçu par les différents photorécepteurs chevauche celui perçu par les pigments photosynthétiques, ce qui permet aux plantes de coordonner leur production d'énergie et leur développement (Lin, 2002). Chez *A. thaliana*, il existe deux grandes classes de photorécepteurs régulant l'initiation de la floraison : les phytochromes et les cryptochromes (Bae et al., 2008).

Les phytochromes³ sont encodés par cinq gènes différents chez *A. thaliana* (PHYA à PHYE). Le produit de ces gènes sont des protéines dimériques constituées de deux apoprotéines identiques qui sont liées de façon covalente à une phytochromobiline⁴. Les phytochromes se présentent selon deux conformations différentes. La forme inactive du phytochrome est la forme Pr. L'absorption de lumière rouge (red : 600-700 nm) par la forme Pr induit sa photo-isomérisation en forme Pfr, la forme active du phytochrome. L'absorption de lumière rouge lointaine (far-red : 700-750 nm) par la forme Pfr induit alors sa photo-isomérisation en forme Pr. Ainsi, l'absorption d'une des deux longueurs d'ondes par la phytochromobiline permet aux phytochromes de passer d'un isomère à l'autre (**Figure 3**).

Il a été montré que, lors du passage de la forme Pr à la forme Pfr, les phytochromes subissent des modifications structurales se traduisant par l'exposition de leurs domaines N-terminal et C-terminal. L'exposition de surfaces d'interaction protéiques au niveau du domaine N-terminal permettrait la transduction du signal qui peut avoir lieu dans le cytosol ou le noyau. L'exposition d'une séquence NLS au niveau du domaine C-terminal permettrait la translocation nucléaire du phytochrome. Une fois dans le noyau, les phytochromes forment des corps nucléaires au sein desquels aurait lieu la transduction du signal, grâce à la présence de plusieurs autres protéines impliquées dans les voies de signalisation dépendantes de la lumière. Le domaine C-terminal des phytochromes présentent également un sous-domaine à activité sérine/thréonine kinase, ce qui est en accord avec une transduction du signal par interaction des phytochromes avec d'autres protéines. Parmi celles-ci, on retrouve par exemple les cryptochromes, PIF3 (PHYTOCHROME INTERACTING FACTOR 3) et COP1 (CONSTITUTIVE PHOTOMORPHOGENIC 1). PIF3 est un facteur de transcription inhibant la photomorphogenèse. Il a été suggéré que les phytochromes activés lient PIF3 afin de le phosphoryler, ce qui activerait son ubiquitination et sa dégradation par le protéasome. Ainsi, les phytochromes peuvent induire les réponses à la lumière en dégradant cet inhibiteur. Il a également été montré que PIF3 et ses homologues interagissent avec différents acteurs de l'horloge circadienne tels que TOC1 et APRR1 (Yamashino *in* Castillon et al., 2007). COP1 est une E3 ubiquitine ligase qui marque pour la dégradation différents facteurs de transcription activant la photomorphogenèse. Il a été suggéré que les phytochromes activés interagissent avec COP1 afin de l'exclure du noyau, ce qui réduit son abondance nucléaire. Ainsi, les phytochromes peuvent induire les réponses à la lumière en excluant ce répresseur.

³ Cette section est inspirée de la revue de Bae et al., 2008

⁴ Phytochromobiline : tétrapyrrole linéaire agissant comme chromophore

Les cryptochromes⁵ sont encodés par deux gènes différents chez *A. thaliana* (CRY1 et CRY2). Le produit de ces gènes sont des protéines très similaires aux photolyases⁶. Ainsi, le domaine N-terminal des cryptochromes présente une grande similarité de séquence avec celui des photolyases. Il est composé de deux chromophores liés de façon non covalente: une FAD dont le rôle est la transmission du signal en aval et une ptérine qui agit comme une antenne percevant la quantité de lumière. Le domaine C-terminal des cryptochromes est quant à lui différent de celui des photolyases (**Figure 4**).

Les cryptochromes sont des protéines essentiellement nucléaires. CRY2 semble constitutivement exprimé dans le noyau, tandis que CRY1 y est plutôt transloqué à l'aide de la séquence NLS présente au niveau de son domaine C-terminal. Une fois dans le noyau, les cryptochromes forment des corps nucléaires au sein desquels aurait lieu la transduction du signal, à l'instar des phytochromes. Le mécanisme d'action des cryptochromes semble assez semblable à celui des photolyases qui catalysent une réaction d'oxydoréduction en réponse à une absorption de lumière. Ainsi, suite à l'excitation de la FAD par les photons, un transfert d'électron pourrait avoir lieu du cryptochrome à une autre molécule impliquée dans la signalisation. Il est également possible que ce transfert d'électron se produise au sein même du cryptochrome, entraînant ainsi un changement conformationnel favorable à de nouvelles interactions protéiques nécessaires à la transduction du signal. Il a également été montré que ces transferts d'électrons pourraient être à la base d'autres modifications biochimiques touchant le cryptochrome (Lin, 2002). En effet, une certaine phosphorylation semble nécessaire au bon fonctionnement des cryptochromes. L'extrémité N-terminale des cryptochromes peut fixer de l'ATP au niveau de la FAD puis, suite à l'excitation du cofacteur par la lumière, transmettre un phosphate à l'extrémité C-terminale riche en résidus sérine. Étant donné que l'extrémité N-terminale des cryptochromes est majoritairement négative, elle va repousser l'extrémité C-terminale phosphorylée, ce qui va engendrer un changement de conformation du cryptochrome pouvant être à la base de nouvelles interactions protéiques nécessaires à la transduction du signal (**Figure 5**). Parmi les protéines interagissant avec les cryptochromes, on retrouve par exemple les phytochromes ainsi que COP1, ce qui suggère que les cryptochromes peuvent induire les réponses à la lumière en régulant négativement ce répresseur de la photomorphogenèse (Lin, 2005).

Il existe une dernière classe de photorécepteurs chez *A. thaliana*. Il s'agit de protéines captant également la lumière bleue grâce à la présence d'un domaine LOV (Light-Oxygen-Voltage) similaire à celui trouvé chez les phototropines. Parmi cette nouvelle classe de photorécepteurs à la lumière bleue, on retrouve les protéines ZEITLUPE (ZTL) et FLAVIN-BINDING KELCH REPEAT F-BOX (FKF1). Ces protéines présentent trois domaines caractéristiques : à l'extrémité N-terminale, un domaine LOV capable de lier un cofacteur (FMN) dont le rôle est la transmission du signal en aval, et à l'extrémité C-terminale un domaine F-box impliqué dans l'ubiquitination de protéines destinées à la dégradation par le protéasome et six répétitions kelch impliquées dans les interactions entre protéines. Les voies de signalisation impliquant ces protéines ne sont pas encore tout à fait connues mais il a été démontré qu'elles étaient capables d'interagir avec d'autres protéines impliquées dans la régulation des voies dépendantes de la photopériode et de les marquer pour la dégradation (Christie, 2007). Ainsi, ZTL module l'accumulation de la protéine TIMING OF CAB1 (TOC1) qui fait partie de l'horloge circadienne, tandis que FKF1 module l'accumulation de la protéine CONSTANS (CO) qui régule les voies de signalisation en aval (Thomas, 2006).

⁵ Cette section est inspirée de la revue de Li et al., 2007

⁶ Photolyase : famille de protéines impliquées dans la réparation de l'ADN endommagé par les rayons ultraviolets (320-390 nm)

Les différents photorécepteurs montrent donc des propriétés structurales et fonctionnelles assez similaires. Cependant, leurs rôles et leurs interactions dans la régulation de l'initiation florale peuvent être assez variables. Ainsi, PHYA capte la lumière rouge de faible intensité (red) afin d'induire la floraison, tandis que les phytochromes B, D et E captent la lumière rouge de haute intensité (far-red) afin d'inhiber la floraison. CRY1 et CRY2 captent la lumière bleue de haute intensité (390-500 nm) afin d'induire la floraison (Lin, 2000). Des études génétiques ont démontré que les phytochromes et les cryptochromes interagissent bien dans la régulation de la photomorphogenèse et de l'initiation de la floraison, ainsi que dans l'entraînement de l'horloge circadienne chez *A. thaliana*. Les cryptochromes fonctionnent également en partenariat avec les phototropines pour la régulation de différents processus induits par la lumière bleue (Li et al., 2007). Étant donné que de nombreux processus physiologiques sont sous le contrôle de différents photorécepteurs, il n'est pas surprenant qu'ils soient simultanément activés par la lumière et que l'action d'un photorécepteur puisse être affectée par la présence des autres photorécepteurs. L'avantage d'une telle multiplicité de photorécepteurs vient de leur capacité à cribler un large spectre de longueurs d'ondes et à fournir ainsi aux plantes un système de signalisation complexe et versatile indispensable pour répondre efficacement aux différents signaux lumineux perçus dans leur environnement (Casal, 2000).

On distingue ainsi deux mécanismes différents par lesquels la lumière affecte la transcription des gènes en aval : soit par une transduction directe du signal d'un photorécepteur aux régulateurs transcriptionnels de ces gènes, soit via la régulation de l'horloge circadienne (Lin, 2002). L'horloge circadienne est composée de trois éléments. Les voies d'entrée qui synchronisent le mécanisme de l'horloge aux cycles quotidiens du jour et de la nuit, un oscillateur central qui génère le périodisme de 24h, et les voies de sortie qui régulent différents processus en aval (Li et al., 2007). De nombreux organismes possèdent ainsi une horloge circadienne qui leur permet d'anticiper les variations de l'environnement et de s'y préparer (Lin, 2002).

Les différents photorécepteurs font partie des voies d'entrée (Li et al., 2007). L'oscillateur central est composé de différentes protéines constituant une boucle de feedback négatif et dont les fonctions convergent toutes vers la régulation de l'expression des gènes en aval, tels que GIGANTEA (GI) et CONSTANS (CO). Enfin, l'activation de ces gènes va mener à l'activation de la transcription du gène FLOWERING LOCUS T (FT), qui lui-même va déclencher le développement floral via l'activation des gènes d'identité florale. Les photorécepteurs induisent donc la transcription des gènes LATE-ELONGATED HYPOCOTYL (LHY) et CIRCADIEN CLOCK-ASSOCIATED 1 (CCA1) dans la première partie du jour (Thomas, 2006). Ces gènes codent pour un facteur de transcription possédant une séquence N-terminale analogue à celle du domaine de liaison à l'ADN des protéines MYB⁷. Leur expression réprime l'expression du gène TOC1, aussi connu comme APRR1 (pseudo-response regulator). Ces gènes répriment également leur propre expression et, lorsque leurs niveaux protéiques diminuent, l'expression de TOC1 augmente. TOC1 favorise alors à son tour l'expression de LHY et de CCA1 afin d'initier un autre cycle. Cette boucle de feedback est capable de réguler la floraison en déterminant le moment de la journée où les autres gènes appartenant à la voie photopériodique, tels que CO, sont exprimés. Le gène CO encoderait une protéine avec deux doigts de zinc très similaires à la boîte-B⁸ liés faiblement à ceux du facteur de transcription GATA. Le gène GI code également pour une protéine nucléaire contrôlant l'expression du gène FT (Mouradov et al., 2002).

⁷ Protéine MYB : protéine nucléaire animale impliquée dans la progression du cycle cellulaire

⁸ Boîte-B : type de doigt de zinc identifié dans les protéines animales et dont on croit qu'il aide aux interactions entre protéines

En effet, l'activation post-traductionnelle de GI et de CO par la lumière mène à l'activation de la transcription de FT (Thomas, 2006). La régulation de FT a lieu dans les feuilles à partir desquelles sa protéine voyage vers l'apex pour initier le développement floral via l'activation des gènes d'identité florale (Boss et al., 2004).

D'autres facteurs sont encore impliqués dans la régulation de la voie photopériodique. Par exemple, différents gènes régulent l'horloge circadienne, tels que les gènes *EARLY-FLOWERING* (ELF) pour lesquels des mutations provoquent une floraison précoce chez *A. thaliana* pendant les jours longs (Mouradov et al., 2002). Les plantes sont également capables de réguler l'initiation de leur floraison en fonction de la qualité et de la quantité de lumière. Ainsi, les phytochromes B, D et E régulent l'expression du gène FT en fonction de la qualité de la lumière via la protéine nucléaire PHYTOCHROME AND FLOWERING TIME 1 (PFT1). Les voies de signalisation impliquées dans la perception de la quantité de lumière sont moins bien connues, mais il est généralement accordé que ce processus est lié à la photosynthèse et à la disponibilité des métabolites, les photorécepteurs étant dans ce cas la chlorophylle et les autres pigments photosynthétiques (Thomas, 2006 – **Figure 6**).

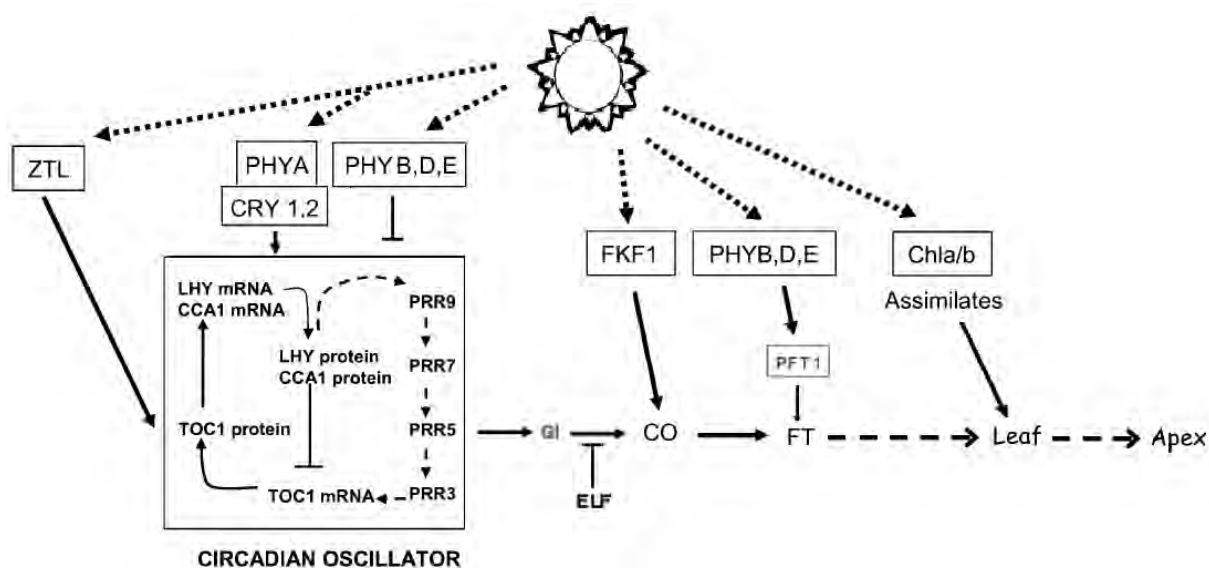


Figure 6 : Voie de floraison dépendante de la photopériode (Thomas, 2006)

Les photorécepteurs ont pour fonction de synchroniser l'horloge circadienne aux cycles quotidiens du jour et de la nuit. Cette boucle de feedback négatif régule ensuite l'expression des gènes GI et CO dont la fonction est d'activer la transcription du gène FT. La protéine FT va finalement initier le développement floral via l'activation des gènes d'identité florale au niveau du méristème apical.

3.2. Voie dépendante de la vernalisation

L'exposition à de basses températures pendant plusieurs semaines accélère généralement la floraison. La sensibilité à ce traitement peut différer entre les espèces. Dans la nature, les variétés nécessitant une vernalisation (annuelles d'hiver) ne fleurissent pas jusqu'à ce qu'elles aient été exposées à des conditions hivernales. Elles germent typiquement pendant l'été, grandissent végétativement pendant l'hiver jusqu'au printemps suivant, et ensuite fleurissent, souvent en réponse à une photopériode appropriée pendant le printemps ou l'été. C'est en contraste avec les annuelles d'été, qui germent et fleurissent pendant le même été sans nécessiter de vernalisation (Mouradov et al., 2002).

Malgré les nombreuses années de recherches consacrées à l'identification des senseurs de la température chez les végétaux, la nature de ces molécules n'est pas encore connue (Penfield, 2008). Cependant, différents acteurs potentiellement impliqués tant dans la perception que dans la transduction du signal au froid ont été identifiés sur base d'études concernant l'acclimatation au froid. Tout comme la vernalisation, ce processus, qui a pour résultat l'acquisition d'une tolérance accrue au froid, nécessite une exposition à de basses températures (Örvar et al., 2000). Cependant, la durée minimum d'exposition au froid nécessaire pour déclencher les processus en aval est bien plus courte dans le phénomène d'acclimatation au froid (~ 1-2 jours) que pour la vernalisation (~ 3 semaines). Cela assure que les plantes fleurissent seulement après la fin de l'hiver et non pas suite aux fluctuations de température rencontrées pendant l'automne.

Les découvertes effectuées au cours des quelques dernières années suggèrent que le premier senseur d'une variation de température serait la membrane plasmique. Ainsi, les observations effectuées par l'équipe d'Örvar leur ont permis de mettre sur pied une théorie pouvant expliquer le mécanisme de perception du froid chez les plantes. La rigidification membranaire induite par le froid déclencherait une réorganisation du cytosquelette physiquement lié aux membranes plasmiques. Cette réorganisation toucherait principalement les microfilaments d'actine composant le cytosquelette. Il existe, dans les cellules animales, des micro-domaines membranaires impliqués dans la signalisation transmembranaire et le remodelage du cytosquelette en réponse à la température : les lipid rafts. Dans les cellules végétales, de précédentes études suggérèrent que la rigidification membranaire et la réorganisation des microfilaments d'actine induites par le froid auraient également lieu au sein de micro-domaines spécifiques de la membrane plasmique (Murata *in* Örvar et al., 2000). Ces données constituaient ainsi les premières preuves d'une implication du cytosquelette dans la transduction du signal menant à la régulation de l'expression génique chez les végétaux, les expériences proposant que les forces physiques au sein du cytosquelette soient traduites en signaux biochimiques au sein de la cellule. En effet, leurs expériences montrent que ces microfilaments d'actine semblent directement ou indirectement reliés à des canaux calciques leur permettant de réguler l'ouverture et la fermeture de ceux-ci. La réorganisation des microfilaments d'actine induite par le froid dissiperait les forces de tension gardant les canaux calciques fermés, ce qui aurait donc pour résultat l'ouverture des canaux et l'augmentation de la concentration intracellulaire en calcium. Cet influx de calcium induit par le froid semble donc constituer un composant intrinsèque de la transduction du signal impliquée dans l'acclimatation au froid. En effet, cette variation de la concentration intracellulaire en calcium peut être ressentie par des protéines kinases dépendantes du calcium (CDPK) qui sont capables d'activer un ensemble d'autres protéines kinases, déclenchant ainsi une cascade aboutissant finalement à la régulation de gènes spécifiques pour l'acclimatation au froid (Viswanathan et al., 2002). On peut donc imaginer que la perception du froid à la base du processus de vernalisation soit similaire à celui décrit pour l'acclimatation au froid, bien qu'aucune preuve appuyant cette théorie n'ait encore vu le jour.

Par contre, il a été démontré que les gènes qui contrôlent la vernalisation et l'acclimatation au froid chez *A. thaliana* sont bien indépendants (Chandler in Smallwood et al., 2002). Le produit d'un gène affectant ces deux processus a néanmoins été mis en évidence (Ishitani, 1998 in Smallwood et al., 2002). Il s'agit du gène HOS1 (High expression of Osmotic Stress-regulated gene expression 1) qui encode une protéine RING finger dont la fonction est proche de celle de la protéine COP1, c-à-d marquer pour la dégradation différents facteurs de transcription activant la photomorphogenèse et activer ainsi la floraison. De plus, tout comme COP1, il a été montré que la protéine HOS1 subit une compartimentation régulée par le froid entre le cytosol et le noyau, pouvant ainsi représenter une connexion entre les événements cytoplasmiques induits par le froid et la transcription de gènes spécifiques dans le noyau (Lee, 2001 in Smallwood et al., 2002).

Les études génétiques menées sur *A. thaliana* ont permis d'identifier un ensemble de gènes responsables de la régulation du processus de vernalisation au sein de la plante. La vernalisation fournit aux plantes la compétence à fleurir en inhibant les gènes répresseurs de la floraison tels que FLOWERING LOCUS C (FLC). Le gène FLC encode un facteur de transcription « MADS box » dont l'expression est régulée positivement par le produit du gène FRIGIDA (FRI). La fonction biochimique de la protéine FRI reste inconnue mais les domaines super-enroulés qu'elle présente suggèrent une potentielle interaction avec d'autres protéines ou avec des acides nucléiques (Boss et al., 2004).

Il a été montré que la répression de FLC induite par la vernalisation est maintenue de manière stable pendant la mitose. Ainsi, une fois que l'expression de FLC est réduite, elle reste basse même dans les organes végétaux formés après l'exposition au froid. Similairement, l'effet de la vernalisation est remis à zéro pendant la méiose (Mouradov et al., 2002). Le niveau d'expression de FLC est donc restauré dans la descendance des plantes vernalisées afin d'assurer que la vernalisation soit encore nécessaire pour la floraison de la génération suivante (Boss et al., 2004). Cette « mémoire de l'hiver » montre donc des caractéristiques similaires avec le contrôle épigénétique de l'expression génique (Mouradov et al., 2002).

Il a ainsi été montré que la vernalisation induit des changements au niveau du « code des histones ». Cette appellation réfère aux différentes modifications covalentes pouvant toucher les histones (méthylation, acétylation, phosphorylation, ubiquitination, ...). Une combinaison spécifique de modifications des histones va influencer la formation d'euchromatine ou d'hétérochromatine et donc favoriser ou inhiber l'expression génique, respectivement. Les modifications spécifiques au sein de la chromatine peuvent donc servir de code pour un switch épigénétique de l'expression génique. En effet, les modifications des histones sont reconnues par des protéines spécifiques associées à la chromatine qui vont recruter d'autres protéines afin de former un complexe d'activation ou de répression capable de maintenir ces modifications stables pendant les divisions cellulaires ultérieures (Sung et al., 2006). La vernalisation induit des changements dans les modifications des histones au niveau du locus FLC (**Figure 7**). En effet, chez les plantes non vernalisées, la chromatine de FLC est très acétylée au niveau des lysines 9 et 14 de l'histone 3 (H3K9 – H3K14) et très méthylée au niveau H3K4. Ces modifications sont caractéristiques de la conformation active de la chromatine (euchromatine). Par contre, chez les plantes vernalisées, la chromatine de FLC est peu acétylée au niveau H3K9 – H3K14, peu méthylée au niveau H3K4 et très méthylée au niveau H3K9 – H3K27. Ces modifications sont caractéristiques de la conformation inactive de la chromatine (hétérochromatine) (Schmitz et al., 2007). Il est probable que d'autres modifications dans les histones de la chromatine de FLC soient encore découvertes ultérieurement. Les protéines reconnaissant ces modifications présentent une région homologue conservée entre les plantes et les animaux appelée « chromodomaine ». Celui-ci leur permet donc de reconnaître les histones modifiés et de s'y lier afin de maintenir

l'activation ou la répression des gènes cibles. Chez *A. thaliana*, la protéine HP1 (Heterochromatin Protein 1) se lie à l'hétérochromatine de FLC au niveau H3K9, ce qui rend FLC inaccessible à la machinerie de transcription maintenant donc sa répression stable suite à la vernalisation (Sung et al., 2005).

Chez *A. thaliana*, la signalisation induite par l'exposition au froid semble aboutir à l'expression du gène VERNALIZATION-INSENSITIVE 3 (VIN3), un répresseur de FLC. Il a été montré que ce gène est seulement exprimé suite à une exposition au froid suffisante et que sa cinétique d'induction est corrélée avec la durée d'exposition au froid (Sung et al., 2005). Ainsi, la répression de FLC serait bien quantitative, c-à-d que FLC pourrait être partiellement réprimé par une durée d'exposition au froid plus courte que celle optimale pour la vernalisation suite à une moindre expression du gène répresseur VIN3 (Sung et al., 2006). Le gène VIN3 encode une protéine contenant un domaine PHD (Plant Homeo Domain) pouvant intervenir dans des interactions avec des acides aminés et/ou des acides nucléiques. Il a également été montré que le domaine PHD est capable de lier des phosphoinositides. Ces phospholipides composant la membrane plasmique sont connus pour être le substrat d'un grand nombre de kinases impliquées dans la signalisation cellulaire. VIN3 pourrait donc représenter un récepteur nucléaire de ces phosphoinositides capable de percevoir les modifications ayant lieu au sein de la membrane plasmique suite à une exposition au froid (Sung et al., 2004). Les protéines présentant un domaine PHD sont généralement trouvées dans les complexes protéiques impliqués dans le remodelage de la chromatine (Sung et al., 2005), ce qui est en accord avec le contrôle épigénétique du gène FLC expliqué ci-dessus. Il a d'ailleurs été montré que la protéine VIN3 était nécessaire au recrutement des autres protéines composant les complexes protéiques régulant la chromatine au locus FLC, telles que VRN1 et VRN2. Ainsi, les gènes VERNALIZATION (VRN) permettent de maintenir la répression du gène FLC suite à la vernalisation. Le gène VRN2 encode une protéine homologue à celles du groupe Polycomb retrouvées chez la Drosophile. Ce sont des répresseurs de la transcription qui modifient la structure de la chromatine de telle façon que les facteurs de transcription ne peuvent pas lier les promoteurs sur l'ADN (Mouradov et al., 2002). VRN1 encode une protéine présentant deux domaines B3 potentiels liant spécifiquement l'ADN végétal, in vitro. Cependant, il n'a pas encore été démontré que VRN1 se lie à FLC, in vivo (Boss et al., 2004). Par contre, il a été montré que les produits de ces deux gènes peuvent interagir entre eux et avec VIN3 dans la régulation de la structure de la chromatine de FLC. Un gène similaire à VIN3 (VIN3-LIKE1, VIL1) a également été identifié comme interagissant avec VIN3 au niveau de son extrémité C-terminale, son extrémité N-terminale présentant quant à elle un domaine PHD semblable à celui de la protéine VIN3.

D'autres régulateurs négatifs de FLC ont également été trouvés tels que les gènes FRIGIDA-ESSENTIAL-1 (FES1) et SUPPRESSOR OF FRIGIDA-4 (SUF4). Les fonctions précises de ces gènes ne sont pas connues mais leur implication dans la régulation de FLC a été clairement démontrée (Schmitz et al., 2007). Il existe également des régulateurs positifs de FLC, tels que les gènes (VIP) qui semblent réguler également de nombreux autres gènes influençant la floraison et d'autres processus de développement (Boss et al., 2004 – **Figure 8**).

La perte de la répression de FLC au cours de la méiose se ferait via la diminution des composants nécessaires au maintien de la répression, telles que les modifications d'histones caractéristiques de l'hétérochromatine de FLC. Une expérience cherchant à montrer une différence d'activité entre le gène FLC dérivé des gamètes maternels et paternels suggère que ce processus de remise à zéro au cours de la méiose pourrait être très proche du mécanisme d'empreinte parentale (Sheldon et al., 2008).

3.3. Voie autonome

La voie autonome a été mise en évidence suite à la découverte d'un groupe de mutants sensibles à la vernalisation et fleurissant tard dans n'importe quelle condition de photopériode. Il s'est avéré par la suite que les gènes impliqués dans cette voie autonome ont pour fonction de réguler négativement l'expression de FLC, agissant ainsi en parallèle avec le processus de vernalisation (Boss et al., 2004). Deux catégories de protéines encodées par les gènes de la voie autonome peuvent être distinguées : des protéines de liaison à l'ARN, comprenant les protéines FCA, FY, FPA, FLK et LD, et des protéines impliquées dans le remodelage de la chromatine, telles que les protéines FLD, FVE et REF6. Il a ainsi été montré que ces différents acteurs de la voie autonome ne fonctionnent pas selon une hiérarchie linéaire mais plutôt comme une série de sous-groupes inhibant un même gène (FLC) par différents mécanismes (Schmitz et al., 2007).

Ainsi, FCA et FY forment un premier groupe épistatique. L'épistasie est une interaction entre différents gènes dans laquelle un gène module l'expression d'un autre. Le gène FCA encode une protéine présentant deux domaines N-terminaux de liaison à l'ARN et un motif d'interaction protéique WW à l'extrémité C-terminale (Boss et al., 2004). L'ARNm de FCA peut être processé à deux positions différentes, ce qui donne lieu à quatre transcrits différents (**Figure 9**). Les différents processing au niveau de l'intron 3 donnent trois transcrits différents. L'intron 3 entier correspond au transcrit α et à moins de 1 % des ARNm produits à partir du gène FCA. Un clivage prématuré et une polyadénylation de l'intron 3 donne le transcrit β (55 %) tandis que l'excision totale de l'intron 3 donne le transcrit γ (35 %). L'excision de l'intron 13 produit quant à elle le transcrit δ (10 %). Le transcrit γ est le seul à encoder la protéine FCA active dans le contrôle de la floraison. Ce processing différentiel de l'ARNm de FCA limite le niveau d'expression du gène dans la plante, tant spatialement que temporellement, ce qui peut limiter l'initiation de la floraison (Macknight et al., 2002). Le gène FY encode une protéine présentant des répétitions WD et deux motifs d'interactions protéiques PPLP à l'extrémité C-terminale. Il a été mis en évidence que la protéine FY d'*A. thaliana* est homologue à un facteur essentiel de polyadénylation impliqué dans le processing de l'ARNm chez les levures. Il a également été montré que les protéines FCA et FY interagissent entre elles par l'intermédiaire de leurs domaines C-terminaux. Il semble que cette interaction soit indispensable pour favoriser la polyadénylation de l'ARNm de FCA ou de FLC via FY, menant donc soit à une diminution du transcrit actif de FCA (ce qui inhibe la floraison) soit à une diminution de l'ARNm de FLC afin d'activer la floraison (Marquardt et al., 2006).

FPA et FVE forment un second groupe épistatique. Le gène FPA encode également une protéine de liaison à l'ARN tandis que le gène FVE encode une protéine présentant des répétitions WD. La manière par laquelle ces deux protéines pourraient interagir entre elles n'a pas encore été établie (Boss et al., 2004).

Par contre, il a été montré que les protéines FVE, FLOWERING LOCUS D (FLD) et RELATIVE OF EARLY FLOWERING 6 (REF6) sont impliquées dans le remodelage de la chromatine à la base de la régulation épigénétique du gène FLC. En effet, ces protéines feraient partie des complexes régulant l'acétylation et la méthylation des histones menant à la formation d'hétérochromatine au niveau du locus FLC (Marquardt et al., 2006).

Les gènes FLOWERING LOCUS K (FLK) et LUMINIDEPENDENS (LD) encodent tous les deux des protéines présentant des domaines de liaison à l'ARN. La protéine LD serait une protéine nucléaire de type « homéobox » connue pour être également impliquée dans le processing de l'ARNm (Boss et al., 2004 – **Figure 10**).

3.4. Voie dépendante de l'acide gibbérellique

L'acide gibbérellique (GA) est une hormone végétale impliquée dans de nombreux processus essentiels au développement de la plante, tels que l'initiation de la floraison chez *A. thaliana* (Schwechheimer, 2008). Il a en effet été montré que l'acide gibbérellique est indispensable au développement des étamines et des pétales lors de la floraison d'*A. thaliana*. L'acide gibbérellique est un diterpène produit par une voie de synthèse complexe impliquant de nombreuses enzymes. Un grand nombre d'acides gibbérelliques différents ont été identifiés chez les plantes, parmi lesquels une petite partie est biologiquement active, les autres étant des intermédiaires biosynthétiques ou cataboliques. Il a ainsi été montré que les protéines encodées par les gènes GA_1 , GA_4 et GA_5 sont des enzymes impliquées dans la synthèse d'acides gibbérelliques biologiquement actifs dans le contrôle de l'initiation de la floraison (Olszewski et al., 2002).

Parmi les facteurs régulant positivement la voie de signalisation de l'acide gibbérellique, on retrouve les protéines GAMYB qui sont des facteurs de transcription MYB induits par l'acide gibbérellique. Il a par exemple été mis en évidence que, sous l'influence de l'acide gibbérellique, la protéine AtMYB33 (*A. thaliana*) peut se lier à une séquence spécifique du promoteur du gène LEAFY (LFY) afin d'induire la floraison (Olszewski et al., 2002).

Parmi les facteurs régulant négativement la voie de signalisation de l'acide gibbérellique, on peut citer les protéines DELLA, telles que GIBBERELLIC ACID INSENSITIVE (GAI), REPRESSOR OF GA1-3 (RGA) et RGA-LIKE 1 (RGL1). Ces protéines nucléaires présentent un domaine carboxy-terminal (RVER) ainsi qu'un motif central (VHYNP) très conservés au sein de la famille de protéines GRAS dont elles font partie, mais également un domaine amino-terminal spécifique (DELLA). Il a été montré que ces régulateurs de la transcription sont capables d'inhiber les voies de signalisation de l'acide gibbérellique en absence de GA actif. Une fois présent, l'acide gibbérellique annule cette inhibition en orientant ces répresseurs vers la dégradation. Il est apparu que le domaine DELLA de ces protéines leur est indispensable pour percevoir la présence de l'acide gibbérellique, ainsi que pour être dégradé. En effet, le récepteur végétal de l'acide gibbérellique (GID1) est une protéine nucléaire soluble qui, une fois lié à l'acide gibbérellique, est capable d'interagir avec le domaine DELLA des protéines du même nom afin d'induire leur dégradation (Schwechheimer, 2008 – **Figure 11**).

Les voies de signalisation de l'acide gibbérellique sont donc fortement régulées par les taux de production et de dégradation de l'acide gibbérellique lui-même (**Figure 12**). Il a également été montré que les stimuli environnementaux, tels que la lumière et la température, pouvaient modifier les concentrations et/ou la réceptivité de la plante à l'acide gibbérellique. En interagissant ainsi avec les autres voies de signalisation régulant le développement de la plante, le métabolisme et les voies de signalisation de l'acide gibbérellique forment un réseau de régulation très complexe (Olszewski et al., 2002). Il a d'ailleurs été montré que, outre la régulation du gène LFY, l'acide gibbérellique est capable de réguler positivement les deux autres gènes « intégrateurs » des voies de floraison, c-à-d les gènes FLOWERING LOCUS T (FT) et SUPPRESSOR OF OVEREXPRESSION OF CO1 (SOC1), afin d'induire l'initiation de la floraison (Boss et al., 2004).

3.5. Intégration des différentes voies de floraison

Les différentes voies décrites ci-dessus constituent la première étape de la régulation génétique du développement floral. Les voies de signalisation ainsi induites, suite aux divers stimuli environnementaux et endogènes perçus par la plante, vont ensuite converger vers la régulation des mêmes gènes en aval. Ces gènes « intégrateurs » ont alors pour fonction de déclencher l'expression des gènes d'identité au niveau du méristème floral qui, à leur tour, vont induire l'expression des gènes d'identité au niveau des organes floraux afin de définir les différentes parties de la fleur (**Figure 13**). Cette régulation génétique complexe permet donc de contrôler le passage de la phase végétative à la phase reproductive au sein de la plante (Jack, 2004).

Les stimuli environnementaux contrôlant la floraison sont perçus par divers organes au niveau de la plante et mènent finalement à l'initiation de la floraison au niveau du méristème apical. L'hypothèse d'un stimulus floral, nommé florigène, pouvant circuler des organes de perception jusqu'au méristème apical a donc été émise. L'identité de ce signal est restée incertaine pendant de nombreuses années, mais les recherches ont finalement abouti à l'identification de la protéine FT (Corbesier et al., 2007). A l'inverse de la plupart des gènes de floraison, qui sont des facteurs de transcription, la protéine FT ne présente pas de domaines fonctionnels spécifiques ni de fonction biochimique évidente. Néanmoins, elle a été apparentée, chez les mammifères, aux protéines associées à la membrane de type PEBP (Phosphatidyl-Ethanolamines Binding Protein) dont l'homologue humain est identique aux protéines inhibitrices des Raf kinases (Putterill, 2001). Ces données suggèrent donc une fonction potentiellement importante de la protéine FT dans la régulation des voies de signalisation. Pour rappel, sous l'influence de longues photopériodes, la protéine CO active la transcription du gène FT au niveau du tissu vasculaire des feuilles. La protéine FT interagit par la suite avec le facteur de transcription FD au niveau du méristème apical afin d'activer le gène d'identité du méristème floral APETALA 1 (AP1). La protéine FT serait donc capable de voyager librement à travers les tissus végétaux via le système vasculaire (phloème). Sa petite taille (23kDa) est d'ailleurs en accord avec un potentiel transport symplastique à partir des feuilles vers le méristème apical où la protéine FT va réguler positivement l'initiation florale (Corbesier et al., 2007).

Parmi les gènes intégrateurs, outre le gène FT qui intègre les signaux de floraison provenant de la voie dépendante de la photopériode, on peut citer le gène FLC. Celui-ci code pour un facteur de transcription « MADS box » réprimant la floraison qui est sous le contrôle des voies dépendantes de la vernalisation et autonome. Les deux autres gènes intégrateurs se trouvent en aval des gènes FT et FLC. Il s'agit des gènes SUPPRESSOR OF OVEREXPRESSION OF CONSTANS (SOC1) et LEAFY qui codent tous les deux pour des facteurs de transcription. Il a été montré que le gène SOC1 est régulé positivement par le gène CO et la voie dépendante de l'acide gibbérellique, alors qu'il est régulé négativement par le gène FLC. Le gène LFY est quant à lui régulé positivement par le gène SOC1 ainsi que par la voie dépendante de l'acide gibbérellique (Jack, 2004).

Le gène LFY est à la fois un gène intégrateur et un gène d'identité du méristème floral. Les gènes d'identité du méristème floral, tels que LFY et AP1, encodent également des facteurs de transcription qui ont pour fonction de spécifier l'identité des méristèmes latéraux afin qu'ils se développent en fleurs plutôt qu'en feuilles ou en pousses. Il a été montré que le gène AP1 est à la fois régulé par le gène LFY et par la voie dépendante de la photopériode via le gène FT (Jack, 2004).

Les gènes d'identité du méristème floral ont pour fonction principale de réguler positivement les gènes d'identité des organes floraux. Ces derniers sont également appelés gènes ABC, sur base du modèle d'interaction génique qu'ils présentent pour spécifier les différentes parties de la fleur. Comme chez la plupart des angiospermes, la fleur d'*A. thaliana* est constituée de quatre organes floraux différents organisés en une série de spirales concentriques. De l'extérieur vers l'intérieur, la fleur présente des sépales déterminés par la classe de gènes A (APETALA 1 et 2), des pétales déterminés par les classe de gènes A et B (APETALA3 et PISTILLATA), des étamines déterminés par les classe de gènes B et C (AGAMOUS) et des carpelles déterminés par la classe de gènes C. Une révision du modèle ABC a par la suite donné naissance au modèle ABCE, la classe de gènes E étant représentée par les trois gènes SEPALLATA (SEP) qui agissent de manière redondante dans la détermination de l'identité des pétales, des étamines et des carpelles (Jack, 2004 – **Figure 14**).

La plupart des gènes cités ci-dessus (FLC, SOC1, AP1, AP3, PI, AG, SEP) sont des facteurs de transcription de type « MADS box » qui se lient généralement à l'ADN sous forme d'homo- ou d'hétérodimères. Des assemblages de protéines « MADS box » plus complexes (tétramères) ont également été mis en évidence dans la détermination de l'identité des organes floraux. Les voies d'intégration décrites ici semblent donc largement dépendantes de la liaison de ces facteurs de transcription à leur promoteur cible, et l'existence de tels complexes constitués tant d'activateurs que de répresseurs fournit à la plante un système de contrôle de la floraison finement régulé en réponse aux différents stimuli perçus (Jack, 2004).

Cette description des différentes voies régulant la floraison chez *A. thaliana* n'est pas exhaustive. La complexité de ces voies de signalisation et d'intégration entraîne la découverte continue de nouveaux facteurs environnementaux régulant la floraison, ainsi que de nouveaux régulateurs génétiques impliquant davantage d'interactions, reflétant ainsi la plasticité des plantes à s'adapter à leur environnement et aux variations saisonnières. Ces données suggèrent également qu'il existe un grand nombre de gènes candidats à étudier. Dans l'étude de la résistance à la montaison chez la chicorée, les gènes impliqués dans les voies dépendantes de la photopériode, de la vernalisation et de la voie autonome, ainsi que les principaux gènes intégrateurs, ont particulièrement été ciblés. Le but final est la sélection de plants de chicorée fleurissant au bon moment, c-à-d la deuxième année. Habituellement, l'activation des gènes FT et SOC1 par CO est empêchée par FLC pendant le premier été. L'expression de FLC est ensuite réduite pendant l'hiver, de telle façon que CO peut activer les gènes FT et SOC1 pendant l'été suivant. La plante ne doit donc pas être exposée au froid durant la première année afin qu'elle ne puisse pas monter au cours de la première saison. Cependant, étant donné que la chicorée est plantée en avril, elle peut être exposée à des températures fraîches, réduisant ainsi l'expression de FLC le premier été, ce qui active finalement la floraison. La plante optimale devrait donc être moins sensible à la vernalisation, afin d'étendre la saison de croissance et d'augmenter le rendement à la récolte via une plantation précoce.

4. Polymorphisme des gènes candidats

Après avoir sélectionné des gènes candidats, il est nécessaire de rechercher leur polymorphisme génétique qui pourra, par la suite, être utilisé pour la conception de marqueurs moléculaires.

4.1. Polymorphisme et marqueurs moléculaires

Un marqueur moléculaire est un locus polymorphe permettant de caractériser le génome d'un organisme (De Vienne, 1998). En effet, les individus au sein d'une même espèce présentent généralement des séquences génomiques conservées, mais également des séquences variables d'un individu à l'autre. Ce polymorphisme permet donc la mise en évidence de différents allèles au sein de la population étudiée. L'avantage des marqueurs moléculaires est qu'ils sont stables au sein des différents organismes, peu importe leur stade de développement au moment de l'analyse. Les marqueurs moléculaires permettent également de réaliser de la cartographie génétique en positionnant les loci polymorphes par rapport à des loci connus auxquels ils sont génétiquement liés, ou de la génétique d'association en déterminant l'impact des loci polymorphes sur les variations phénotypiques observées (Pflieger et al., 2001).

Le polymorphisme au sein du génome peut être de trois types différents : variation du nombre de répétitions dans les microsatellites, insertion/délétion ou mutation ponctuelle dans une séquence. Dans ce travail, les mutations ponctuelles de type SNP (Single Nucleotide Polymorphism) ont été utilisées pour la mise en évidence du polymorphisme génétique des gènes candidats sélectionnés. Les SNP apparaissent suite à l'insertion, la délétion ou la substitution d'un nucléotide qui peut soit être une transition, c-à-d un échange entre deux purines ($A \leftrightarrow G$) ou deux pyrimidines ($C \leftrightarrow T$), ou une transversion, c-à-d un échange entre une purine et une pyrimidine (De Vienne, 1998). Les transitions nucléotidiques sont plus fréquemment observées que les transversions, ce qui pourrait s'expliquer par la plus haute fréquence de mutations touchant les sites CpG (cytosine suivie d'une guanine) une fois qu'ils sont méthylés pour la régulation épigénétique. En effet, la déamination spontanée de la cytosine forme de l'uracile qui est reconnu et remplacé par une autre cytosine par l'action des enzymes de réparation de l'ADN. Par contre, la déamination spontanée de la cytosine méthylée (5-méthylcytosine) forme de la thymine qui n'est pas reconnue comme une erreur par les enzymes de réparation. Cette conversion de cytosine en thymine persiste donc au sein du génome, donnant lieu à une transition stable et fréquente (Scarano *in* De Vienne, 1998).

Les SNP sont des marqueurs co-dominants (c-à-d qu'ils permettent de distinguer les individus homo- ou hétérozygotes représentant la population étudiée) et bialléliques (c-à-d que le polymorphisme observé au niveau d'un nucléotide ne présente que deux alternatives). La nature de la base modifiée est généralement utilisée pour classer les différents types de SNP. En effet, le changement d'un nucléotide par un autre peut avoir un impact plus ou moins important selon sa localisation ou la modification qu'il entraîne au niveau de la séquence protéique. Ainsi, on distingue les SNP non codants localisés dans les introns et les régions 5' et 3'-UTR et les SNP codants localisés dans les exons. Les SNP peuvent également être subdivisés en fonction des mutations induites. Une mutation silencieuse est un changement nucléotidique qui, grâce à la redondance du code génétique, n'entraîne pas de modification de l'acide aminé et n'a donc aucun impact sur le phénotype. Une mutation non-sens est un changement nucléotidique entraînant la conversion d'un codon codant en codon stop, ce qui induit la production d'une protéine tronquée. Une mutation faux-sens est un changement nucléotidique entraînant une modification de l'acide aminé, ce qui peut avoir une répercussion différente en fonction de sa localisation. Un changement d'acide aminé dans la séquence

codante peut altérer la structure, et donc la fonction, de la protéine produite, tandis qu'un changement d'acide aminé dans la séquence non codante peut modifier l'affinité d'un promoteur pour les facteurs de transcription ou encore l'épissage alternatif des introns, ce qui a un impact important dans la régulation de l'expression génique. Enfin, l'insertion ou la délétion d'un nucléotide va décaler le cadre de lecture de la séquence, ce qui entraîne généralement la production d'une protéine tronquée par l'apparition de codons stop au sein de la séquence nucléotidique.

La fréquence des SNP au sein du génome d'un organisme peut varier en fonction de l'espèce, mais elle est en moyenne de 1 SNP/100 pb. L'avantage d'étudier le polymorphisme de type SNP plutôt qu'un autre réside dans le fait que ces mutations ponctuelles constituent la majorité du polymorphisme détecté au sein des génomes et qu'elles y sont réparties de façon homogène. Cela permet par exemple de créer des cartes génétiques plus complètes, augmentant ainsi la probabilité de découvrir un SNP génétiquement lié à un locus connu ou statistiquement associé aux variations phénotypiques observées, qu'il en soit la cause même en étant localisé au sein du gène responsable du phénotype ou qu'il y soit juste lié. A l'échelle d'un génome entier, la majorité des SNP n'a cependant aucun impact direct sur le phénotype.

4.2. Recherche de SNP

Il existe un grand nombre de techniques permettant de mettre en évidence du polymorphisme au sein d'un génome. Ces techniques nécessitent généralement une connaissance préalable de la séquence à analyser, permettant ainsi de cibler des loci spécifiques et/ou de génotyper le polymorphisme détecté. Le procédé de détection peut être basé sur la mise en évidence de sites de restriction différents (RFLP et AFLP), de stabilité (DGGE et DHPLC) ou de conformation différente (SSCP), ou encore de sites d'hybridation différents (ISSR et sondes oligonucléotidiques). Les résultats peuvent également être visualisés en phase liquide (gel) ou sur support solide (microarray). Ces techniques sont généralement assez fiables et relativement rapides pour la recherche de polymorphisme, quoique présentant toujours certains avantages et/ou inconvénients selon les résultats recherchés (Kwok et al., 2003).

Une méthode alternative a été sélectionnée pour la caractérisation du polymorphisme de nos gènes candidats. Il s'agit de l'EcoTILLING (**Figure 15**), une adaptation de la technique du TILLING (Targeting Induced Local Lesions in Genomes) qui permet la détection de polymorphisme dans les populations naturelles. Pour détecter un SNP dans un gène d'intérêt par EcoTILLING, il faut réaliser une PCR avec des amorces spécifiques du gène qui sont marquées avec des fluorochromes. Après amplification, les échantillons sont dénaturés puis réhybridés, induisant ainsi la formation d'hétéroduplexes entre les brins d'ADN polymorphes. Les mésappariements à la position des SNP sont ensuite partiellement clivés par une endonucléase spécifique de simple brin (Endo-I). Les fragments fluorescents résultant de cette digestion sont finalement séparés par électrophorèse, ce qui permet de localiser et de dénombrer les SNP (Till et al., 2006). En effet, la position de la mutation, et donc du clivage, va influencer la vitesse de migration des fragments fluorescents. Un SNP localisé près de l'extrémité marquée va engendrer un petit fragment fluorescent ayant une vitesse de migration rapide, tandis qu'un SNP éloigné de l'extrémité marquée va induire un grand fragment fluorescent présentant une vitesse de migration plus lente (Comai et al., 2004). Au laboratoire, les fragments sont séparés par électrophorèse capillaire à l'aide d'un analyseur de taille de fragments, ce qui permet de numériser les résultats. Ceux-ci apparaissent ainsi sous formes de pics, chacun d'entre eux correspondant à un SNP (*voir figure 21*). L'utilisation de cette technique est plus rapide et les résultats plus faciles à manipuler par rapport à une séparation sur gel d'électrophorèse.

L'EcoTILLING représente donc une technique permettant de détecter rapidement un grand nombre de SNP dans diverses séquences d'intérêt, d'une manière reproductible et relativement abordable d'un point de vue financier. De plus, ce procédé ne nécessite pas de connaissance préalable de la séquence à analyser, excepté pour la conception des amorces marquées spécifiques du gène. Si l'EcoTILLING ne permet pas de génotyper les SNP détectés, il constitue néanmoins une analyse préalable de la localisation et du nombre de SNP permettant de sélectionner facilement les séquences qui seraient les plus intéressantes à génotyper, du point de vue du polymorphisme.

4.3. Génotypage de SNP⁹

Le séquençage est sans doute la méthode la plus simple pour déterminer la nature des SNP détectés en EcoTILLING, mais elle reste aussi la plus coûteuse. Il existe bien d'autres techniques permettant de génotyper les SNP et, dans cette optique, il reste à mettre au point au laboratoire un système permettant de multiplexer le génotypage de nos gènes candidats. Ce système sera une variante du Golden Gate Assay qui combine trois techniques différentes dont le principe général consiste à discriminer les deux allèles existants, en présence d'un SNP au locus étudié, via l'utilisation d'oligonucléotides spécifiques d'allèles. Ainsi, seul l'oligonucléotide parfaitement complémentaire à la séquence analysée peut s'y lier de façon stable.

L'hybridation d'oligonucléotide utilise deux sondes spécifiques d'allèles présentant la base polymorphe généralement au milieu de la séquence oligonucléotidique. Le génotypage du SNP est uniquement basé sur l'hybridation stable de la sonde complémentaire au locus analysé (**Figure 16**).

L'extension d'amorce utilise deux amorces spécifiques d'allèles présentant la base polymorphe à l'extrémité 3' de la séquence oligonucléotidique. La polymérase est capable de prolonger l'oligonucléotide uniquement lors d'une hybridation stable de l'amorce complémentaire au locus analysé (**Figure 17**).

La ligation d'oligonucléotide utilise deux sondes spécifiques d'allèles présentant la base polymorphe à l'extrémité 3' de la séquence oligonucléotidique, ainsi qu'une sonde adjacente spécifique du locus. La ligase est capable d'assembler les oligonucléotides adjacents uniquement lors d'une hybridation stable de la sonde complémentaire au locus analysé (**Figure 18**).

Le système qui doit être mis au point au laboratoire n'utilisera pas la technique de l'extension d'amorces. Dans notre système, la séquence d'ADN à analyser est mise en présence de deux sondes spécifiques d'allèles (ASO) et d'une sonde spécifique de locus (LSO). Les deux ASO présentent la base polymorphe à l'extrémité 3' et une séquence reconnue par des amorces universelles (RV) à l'extrémité 5'. La LSO contient une séquence complémentaire à un ZipCode, ainsi qu'une séquence reconnue par une amorce universelle (FW), à l'extrémité 3'. L'ASO complémentaire et la LSO vont s'hybrider de manière stable à la séquence avant d'être assemblées par la ligase. Le produit de ligation ainsi obtenu est ensuite amplifié par les amorces universelles avant d'être hybridé sur un support solide (microarray) présentant différents ZipCodes. L'utilisation de deux amorces universelles RV différentes marquées par divers fluorochromes permet d'effectuer une PCR marquant différenciellement les deux allèles. La détection de la fluorescence émise sur le support solide permet donc d'identifier spécifiquement le produit PCR obtenu afin de génotyper le SNP (**Figure 19**).

⁹ Cette section est inspirée de l'article de Sobrino et al., 2005

Il existe évidemment d'autres techniques permettant de génotyper les SNP, ainsi que différentes méthodes de détection. S'il n'existe pas vraiment de procédé idéal, une certaine optimisation est néanmoins recherchée, notamment en ce qui concerne la précision, la sensibilité et la reproductibilité de la technique utilisée. Le fait de pouvoir multiplexer les analyses, c-à-d pouvoir détecter en une seule fois un grand nombre de SNP différents au sein de diverses séquences, reste généralement la caractéristique la plus recherchée lors du choix de la technique à utiliser en routine, et ce pour un gain de temps et d'argent. Dans le cas du GoldenGate Assay, l'utilisation de différents modificateurs de mobilité au niveau de la LSO ainsi que de différents fluorochromes au niveau des ASO permettraient de discriminer les produits PCR obtenus par taille et par couleur, et donc de multiplexer les analyses préliminaires effectuées sur séquenceur.

4.4. Utilisation des SNP en génomique¹⁰

Une fois que le polymorphisme de nos gènes candidats sera mis en évidence et que la technique d'analyse à haut débit de ce polymorphisme sera mise au point, des études plus approfondies en génomique structurale (cartographie génétique) et fonctionnelle (études d'association) pourront être réalisées afin de concevoir des marqueurs moléculaires à utiliser en sélection.

La variation phénotypique de la plupart des caractères d'intérêt agricole est influencée par plusieurs gènes, ainsi que par l'interaction de ces différents gènes entre eux et avec l'environnement. L'étude de ces caractères complexes peut être facilitée par l'association d'un haplotype, c-à-d la combinaison d'allèles caractérisant un individu, au phénotype étudié. L'approche la plus souvent utilisée dans ce contexte est l'analyse du déséquilibre de liaison permettant la conception de cartes génétiques. Le déséquilibre de liaison est défini comme une répartition d'allèles non aléatoire au sein de différents loci. Ce phénomène dépend essentiellement de la distance comprise entre ces loci, ainsi que de la fréquence de recombinaison des chromosomes les portant. Ainsi, plus la distance entre deux allèles est importante, plus la probabilité d'observer une recombinaison entre eux est élevée et plus le déséquilibre de liaison entre ces deux allèles est faible. L'idée est que les SNP présents au sein de séquences génétiquement liées voyagent ensemble, et que la mise en évidence d'un de ces SNP permet d'extrapoler la présence des autres qui y sont liés. Cela permet donc de cartographier les SNP détectés dans les gènes candidats grâce au déséquilibre de liaison qu'ils partagent avec d'autres SNP connus. Lors des études d'association, le but est de relier le polymorphisme détecté dans les gènes candidats aux variations phénotypiques observées. Les SNP détectés peuvent en effet être liés à la séquence responsable du phénotype ou, plus rarement, être la cause même du phénotype. L'intérêt d'avoir plusieurs SNP définissant un même haplotype est qu'en présence d'un déséquilibre de liaison dans la séquence d'intérêt, certains des SNP deviennent redondants et seuls quelques-uns d'entre eux sont alors suffisants pour mettre en évidence l'haplotype recherché pouvant être lié au phénotype observé.

Lorsque de telles associations auront pu être mises en évidence, des expériences de phénotypage en champ devront être réalisées afin de valider ces observations. Une fois que ces marqueurs moléculaires seront validés, ils pourront être utilisés pour l'analyse à haut débit de tout individu, devenant donc la base de la sélection assistée par marqueurs dans le domaine de la résistance à la montaison.

¹⁰ Cette section est inspirée de l'article de Zhu et al., 2008

5. Objectifs du mémoire

La chicorée industrielle (*Cichorium intybus* L.) est une plante possédant des racines riches en inuline. Cette plante bisannuelle nécessite une exposition au froid et à de longues photopériodes pour atteindre sa phase reproductive. Dans certaines conditions, la plante monte et fleurit dès la première année, réduisant ainsi les ressources carbonées stockées dans les racines. L'objectif de ce mémoire est la recherche de marqueurs moléculaires pouvant améliorer et accélérer la sélection de plantes résistantes à la montaison.

L'approche gène candidat semble constituer une méthode adaptée à l'étude du processus de floraison qui est régulé par une multiplicité de gènes interagissant entre eux et avec l'environnement. Les voies de signalisation à la base de cette régulation sont maintenant bien connues chez la plante modèle *Arabidopsis thaliana*. La sélection de gènes candidats fonctionnels sera donc basée sur les informations disponibles à propos des quatre voies principales de floraison connues chez *A. thaliana*.

Une fois ces gènes candidats sélectionnés dans les bases de données informatiques, des amorces permettant d'amplifier l'ADNg extrait des feuilles de chicorée seront réalisées. Le clonage orienté de ces produits PCR au sein d'un vecteur se fera ensuite par une étape de restriction-ligation simultanée. La transformation bactérienne par choc thermique permettra alors de sélectionner, par criblage PCR, les colonies bactériennes positives, c-à-d celles contenant bien le vecteur avec l'insert d'intérêt.

La recherche du polymorphisme génétique de type SNP au sein des gènes candidats sera finalement effectuée. La détection des SNP sera réalisée par EcoTILLING et le génotypage des SNP observés sera effectué par séquençage. L'objectif final est la mise au point d'une technique d'analyse à haut débit dérivée de la méthode du Golden Gate Assay. Celle-ci pourra effectivement être utilisée pour des études ultérieures en génomique structurale (cartographie génétique) et fonctionnelle (études d'association), le but final étant la conception de marqueurs moléculaires utilisables en sélection.

1. Extraction de l'ADNg des feuilles de chicorée

Matériels et solutions

- Nucleon Phytopure Genomic DNA Extraction Kit (*Amersham Biosciences* #RPN 8510)
- Azote liquide
- phénol : chloroforme : alcool isoamylique (PCI 25:24:1; *Sigma* #087K1890)
- Chloroforme stocké à -20° C (*Acros Organics*)
- Isopropanol stocké à -20° C (*Sigma*)
- Ethanol 70 % stocké à -20° C (*Sigma*)
- Tris 10 mM pH 8 (*Sigma*)

Protocole

- Broyage des feuilles
 - Maintenir les échantillons congelés pendant toute la procédure en plaçant les eppendorfs contenant les feuilles dans l'azote liquide
 - Broyer les feuilles dans les eppendorfs à l'aide d'un pilon en acier inoxydable jusqu'à obtenir une fine poudre
- Extraction de l'ADN
 - Ajouter à chaque tube 600 µl de réactif 1 puis 200 µl de réactif 2, et mélanger les tubes par inversion afin d'obtenir un mélange homogène
 - Incuber pendant 10' dans un bain à 65° C en mélangeant régulièrement les tubes
 - Laisser refroidir les tubes pendant 10' à température ambiante
 - Ajouter à chaque tube 500 µl de chloroforme stocké à -20° C puis 100 µl de résine Nucleon Phytopure DNA Extraction, et mélanger par inversion afin de précipiter les protéines et les polysaccharides végétaux
 - Placer sur agitateur rotatif pendant 10' à température ambiante
 - Placer sur glace pendant 20'
 - Centrifuger à 1.300 g pendant 10' à température ambiante puis récupérer le surnageant
 - Ajouter à chaque tube 500 µl du mélange PCI 25:24:1 et mélanger par inversion afin de purifier l'ADN
 - Centrifuger à 13.000 rpm pendant 5' à température ambiante (*Biofuge 13, Heraeus*) puis récupérer la phase aqueuse
 - Réaliser de nouveau les deux étapes précédentes
 - Ajouter à chaque tube un volume d'isopropanol stocké à -20° C et mélanger par inversion afin de précipiter l'ADN
 - Centrifuger à 13.000 rpm pendant 10' à température ambiante puis éliminer le surnageant
 - Ajouter à chaque tube un volume d'éthanol 70 % stocké à -20° C afin de laver l'ADN
 - Centrifuger à 13.000 rpm pendant 5' à température ambiante puis éliminer le surnageant
 - Sécher les culots au Speedvac pendant 15'
 - Ajouter à chaque tube 20 µl de Tris 10 mM pH 8 afin de resuspendre l'ADN
 - Placer sur agitateur rotatif overnight à 4° C
 - Stocker les échantillons à -20° C jusqu'à utilisation

2. Design d'amorces et amplification des gènes candidats

➤ Design d'amorces

Les amorces ont été réalisées à l'aide du logiciel **Primer3** (<http://frodo.wi.mit.edu/>) en utilisant les paramètres suivants :

- longueur de l'amorce ≈ 20 nucléotides
- longueur du produit amplifié ≈ 300 -500 pb
- $T_m \approx 50$ -60° C
- % GC ≈ 50 %
- GC clamp ≈ 1 -2

Les propriétés thermodynamiques des amorces (formation de dimères et d'hairpins) ont été analysées à l'aide du logiciel **Vector NTI Advance 10**. Afin de permettre le clonage orienté des produits PCR, les sites de restriction Sall et SpeI ont également été ajoutés aux extrémités des amorces, ainsi que quelques paires de bases supplémentaires permettant d'augmenter l'efficacité de restriction.

➤ Amplification des gènes candidats par PCR (50 μ l)

- 0,3 μ l amorce forward (FW) 100 μ M (*IDT*)
- 0,3 μ l amorce reverse (RV) 100 μ M (*IDT*)
- 1 μ l ADNg (5 ng/ μ l)
- 0,5 μ l DNTP 10 mM (*Fermentas*)
- 0,4 μ l GoTaq Polymerase (*Promega*, 5U/ μ l)
- 10 μ l tampon GoTaq 5x (*Promega*)
- 37,5 μ l eau
- cycle PCR : 4' à 94° C – [45'' à 92° C – 45'' à 55° C – 45'' à 72° C] x 35 – 10' à 72° C – 4° C à l'infini

3. Clonage des gènes candidats

Le plasmide pBluescript II KS+ utilisé pour le clonage des gènes candidats a été produit par maxiprep afin de transformer des bactéries thermocompétentes.

3.1. Préparation de cellules thermocompétentes

Les cellules thermocompétentes ont été produites selon le protocole décrit sur le site **OpenWetWare** (http://openwetware.org/wiki/TOP10_chemically_competent_cells).

Matériels et solutions

- *E. coli* TOP 10F' (*Invitrogen* #C303003)
- Milieu LB liquide (*BD* #240230)
- Tampon CCMB80 (pH 6,4) stocké à 4° C
 - 10 mM KOAc pH7,0
 - 80 mM CaCl₂.2H₂O (*VWR* #1226)
 - 20 mM MnCl₂.2H₂O (*Merck* #5934)
 - 10 mM MgCl₂.6H₂O (*Merck* #1.05833.0250)
 - 10 % glycerol (*Sigma* #G-7757)

Protocole

- Inoculer 400 ml de milieu LB avec 10 ml d'une pré-culture bactérienne lancée overnight et 4 ml de tétracycline (50 µg/ml; *Sigma*)
- Incuber à 180 rpm et 37° C (*NBS*, C24 Incubator Shaker) jusqu'à atteindre une D.O._{590nm} de 0,3
- Aliquoter 200 ml de la culture bactérienne dans quatre tubes de 50 ml placés sur glace et sous hotte à flux laminaire vertical
- Centrifuger à 4.500 rpm pendant 10' à 4° C (*Allegra X-22R*, *Beckman Coulter*) puis éliminer le surnageant
- Ajouter à chaque tube 20 ml de tampon CCMB80 stérile stocké à 4° C
- Resuspendre les culots à l'aide d'une pipette puis pooler en deux tubes de 40 ml
- Incuber pendant 20' sur glace
- Centrifuger à 4.500 rpm pendant 10' à 4° C puis éliminer le surnageant
- Ajouter à chaque tube 2 ml de tampon CCMB80 stérile stocké à 4° C
- Resuspendre les culots à l'aide d'une pipette puis pooler en un seul tube
- Tester la D.O._{590nm} en mélangeant 200 µl de cellules avec 800 µl de milieu LB jusqu'à atteindre 1-1,5
- Incuber pendant 20' sur glace
- Aliquoter par 50 µl en eppendorfs
- Stocker à -80° C jusqu'à utilisation

3.2. Transformation bactérienne par choc thermique

Matériels et solutions

- *E. coli* TOP 10F' thermocompétentes
- pBluescript II KS + (*Stratagene*)
- Milieu LB solide (boîtes de Pétri)
 - LB (*BD* #240230)
 - Agar (*Fluka* #05038)
 - Ampicilline (60 µg/ml; *Sigma* #A0166)
 - X-gal (87,5 µg/ml; *Promega* #V3941)
 - IPTG (41,7 µg/ml)

- Milieu SOC (pH 7,0) stocké à température ambiante
 - 2 % bacto-peptone (*BD* #211677)
 - 0,5 % yeast extract (*BD* #212750)
 - 2,5 mM KCl (*VWR* #1591)
 - 10 mM MgCl₂ (*Merck* #1.05833.0250)
 - 10 mM MgSO₄ (*VWR* #1433)
 - 20 mM glucose (*Riedel-deHaën* #16301) stérilisé sur filtre 0,22 µm à rajouter après autoclavage du milieu

Protocole

- Ajouter 0,5 µg de plasmide aux 50 µl de bactéries thermocompétentes placées sur glace et sous hotte à flux laminaire vertical
- Incuber pendant 30' sur glace
- Effectuer le choc thermique dans un bain à 42° C pendant 30''
- Transférer immédiatement sur glace pendant 2'
- Ajouter immédiatement 250 µl de milieu SOC stérile stocké à température ambiante
- Etaler 150 µl sur milieu LB solide avec X-gal et IPTG
- Incuber overnight à 37° C

3.3. Maxiprep

Matériels et solutions

- QIAGEN Plasmid Maxi Kit (*QIAGEN* #12162)
- *E. coli* TOP 10F' transformées avec pBluescript II KS +
- Milieu LB liquide (*BD* #240230)
- Isopropanol stocké à température ambiante (*Sigma*)
- Ethanol 70 % stocké à température ambiante (*Sigma*)
- Tris 10 mM pH 8

Protocole

- Inoculer 400 ml de milieu LB avec 20 ml d'une pré-culture bactérienne lancée overnight et 240 µl d'ampicilline (60 µg/ml; *Sigma*)
- Incuber à 180 rpm et 37° C (*NBS*, C24 Incubator Shaker) jusqu'à atteindre une D.O._{590nm} de 0,4-0,8
- Aliquoter la culture bactérienne dans deux biberons de 250 ml placés sur glace
- Centrifuger à 6.000 g pendant 15' à 4° C puis éliminer le surnageant
- Ajouter à chaque biberon 5 ml de tampon P1 stocké à 4° C puis 5 ml de tampon P2 stocké à température ambiante et mélanger par inversion
- Incuber pendant 5' exactement à température ambiante puis pooler dans un tube correx stérile de 30 ml
- Ajouter 10 ml de tampon P3 stocké à 4° C et mélanger par inversion
- Incuber sur glace pendant 20'
- Centrifuger à 20.000 g pendant 30' à 4° C puis récupérer le surnageant
- Préparer pendant ce temps la colonne de fixation des plasmides QIAGEN-tip 500 en faisant passer 10 ml de tampon QBT
- Faire passer le surnageant sur la colonne afin de fixer l'ADN

- Faire passer 2 x 30 ml de tampon QC afin de laver la colonne
- Eluer l'ADN avec 15 ml de tampon QF dans un tube correx stérile de 30 ml
- Ajouter 10,5 ml d'isopropanol stocké à température ambiante et mélanger par inversion afin de précipiter l'ADN
- Centrifuger à 15.000 g pendant 30' à 4° C puis éliminer le surnageant
- Ajouter 5 ml d'éthanol 70 % stocké à température ambiante afin de laver l'ADN
- Centrifuger à 15.000 g pendant 10' à 4° C puis éliminer le surnageant
- Sécher le culot au Speedvac pendant 15'
- Ajouter 50-100 µl de Tris 10 mM pH 8 afin de resuspendre l'ADN
- Doser l'ADN par spectrophotométrie (D.O._{260nm}) pour avoir une concentration finale d'environ 1 µg/µl

3.4. Linéarisation du vecteur

- Restriction du vecteur (20 µl)
 - Préparer un mélange d'enzymes de restriction contenant deux fois plus de XhoI (*NEB* #R0146S) que de XbaI (*NEB* #R0145S)
 - 1 µl du mélange d'enzymes (13,3U XhoI + 6,7U XbaI)
 - 2 µl tampon 2 10x (*NEB*)
 - 0,5 µl BSA 100x (*NEB*)
 - 4 µl de vecteur (1 µg/µl)
 - 13 µl eau
 - Incuber pendant 4h30 à 37° C (étuve)
- Déphosphorylation du vecteur linéarisé (50 µl)
 - 20 µl de vecteur digéré
 - 1 µl de phosphatase alcaline (CIAP, 1U/µl; *Promega* #M1821)
 - 5 µl du tampon de la phosphatase alcaline 10x (*Promega*)
 - 23 µl d'eau
 - Incuber pendant 30' à 37° C (thermocycler)
 - Ajouter de nouveau 1 µl de phosphatase alcaline
 - Incuber pendant 30' à 37° C

4. Purification des produits PCR et du vecteur

- Utilisation du mélange phénol : chloroforme : alcool isoamylique (PCI 25:24:1)
 - Ajouter 50 µl de PCI 25:24:1 (*Sigma* #087K1890) aux tubes contenant les produits PCR et le vecteur linéarisé, et mélanger par inversion
 - Centrifuger à 13.000 rpm pendant 5' à 4° C (Biofuge A, *Heraeus*) et récupérer la phase aqueuse
 - Ajouter à chaque tube 10 µl de tampon de charge (*Promega*)
 - Faire migrer les échantillons dans un gel d'agarose 1 % avec 10 µl de Smart Ladder (SL) comme standard de taille

➤ Extraction du gel d'agarose

Les fragments d'ADN ont été découpés du gel et purifiés à l'aide du kit **NucleoSpin Extract II** (Macherey-Nagel #740 609.50).

5. Etape de restriction-ligation

Les étapes de restriction et de ligation des produits PCR ont été réalisées simultanément, selon un ratio de 3:1 (produit PCR/vecteur) dans un volume final de 20 µl :

- 1 µl vecteur linéarisé (50 ng/µl)
- **X** µl produit PCR
- **Y** µl eau
- 0,5 µl SalI-HF (*NEB* #R3138S) = 10U
- 0,5 µl SpeI (*NEB* #R0133S) = 10U
- 0,3 µl BSA 100x (*NEB*)
- 0,3 µl T4 DNA ligase (*NEB* #M0202S) = 6000U
- 2µl de T4 DNA Ligase Reaction Buffer 10X (*NEB* #B0202)
- 1 µl NaCl 1 M
- paramètres : 2h30 à 37° C – 3h à 20° C – 16° C à l'infini

6. Transformation bactérienne par choc thermique (voir section 3.2.)

50 µl de cellules thermocompétentes *E. coli* TOP 10F' ont été transformées avec 3 µl de produit de restriction-ligation selon le protocole expliqué dans la section 3.2.

7. Screening

Les colonies bactériennes transformées ont été criblées par PCR à l'aide des amorces T3 et T7 marquées avec un dye WellRED bleu (D4) ou vert (D3), respectivement.

- 1 µl amorce T7 (FW) 10 µM (*Sigma*)
- 1 µl amorce T3 (RV) 10 µM (*Sigma*)
- 0,5 µl DNTP 10 mM (*Fermentas*)
- 0,4 µl GoTaq Polymerase (*Promega*, 5U/µl)
- 10 µl tampon GoTaq 5x (*Promega*)
- 37,1 µl eau
- cycle PCR : 4' à 94° C – [45'' à 92° C – 45'' à 58° C – 45'' à 72° C] x 35 – 10' à 72° C – 4° C à l'infini

8. EcoTILLING

Matériels et solutions

- Mélange de digestion (11 µl)
 - 1,5 µl Endo-I (10U/µl; *Serial Genetics* #E0515) dilué 50x dans son tampon de dilution
 - 1,5 µl de tampon de réaction 10x
 - 8 µl d'eau milliQ
- Solution stop (5 µl)
 - 2 µl acétate de sodium 3 M pH 5,2 (*Sigma* #S7899)
 - 0,4 µl EDTA 0,5 M pH 8 (*Sigma* #E7889)
 - 1 µl glycogène 10 mM (*Roche* #10901393001)
 - 1,6 µl d'eau milliQ
- Ethanol absolu stocké à -20° C (*Sigma*)
- Ethanol 70 % stocké à -20° C (*Sigma*)
- Sample Loading Solution (SLS, *Beckman Coulter* #608082)
- Standard de taille MapMarker1000 (*BioVentures Inc*)

Protocole

- Dénaturation-hybridation de l'ADN
 - Dénaturer 10 µl de produits PCR pendant 10' à 95° C (thermocycler), puis réhybrider en diminuant graduellement la température (0,5-1° C/minute) jusqu'à 4° C
- Digestion de l'ADN
 - Mélanger 11 µl de mélange de digestion avec 4 µl de chaque échantillon d'ADN hybridé
 - Incuber pendant 2h à 37° C (à l'étuve)
- Précipitation de l'ADN
 - Ajouter à chaque échantillon 5 µl de solution stop et 60 µl d'éthanol absolu stocké à -20° C afin de stopper la réaction
 - Centrifuger à 3.500 rpm pendant 1h à 4° C (ALC PK 120R, *DJB labcare*) et éliminer le surnageant
 - Ajouter à chaque échantillon 200 µl d'éthanol 70 % stocké à -20° C afin de laver l'ADN
 - Centrifuger à 3.500 rpm pendant 15' à 4° C et éliminer le surnageant
 - Réaliser de nouveau les deux étapes précédentes
 - Sécher les culots au Speedvac pendant 15'
 - Mélanger pendant ce temps 340 µl de SLS avec 2 µl de standard de taille
 - Ajouter à chaque échantillon 40 µl de SLS contenant du standard de taille afin de resuspendre l'ADN
 - Séparer les fragments sur un séquenceur à capillaires (CEQ 8000 fragment analyser, *Beckman Coulter*)

9. Miniprep

Matériels et solutions

- Milieu LB liquide (*BD #240230*)
- Glycérol 40 %
- Isopropanol stocké à -20° C (*Sigma*)
- Ethanol 70 % stocké à -20° C (*Sigma*)
- Tris 10 mM pH 8
- Tampon P1 :
 - 100 µg/ml Rnase A (*Sigma*)
 - 50 mM Tris HCl pH 8 (*Sigma*)
 - 10 mM EDTA (*Sigma*)
- Tampon P2 :
 - 200 mM de NaOH
 - 1 % SDS
- Tampon P3 : 3 M de KAc pH 5,5

Protocole

- Inoculer 4 ml de milieu LB avec la colonie bactérienne d'intérêt et 12 µl d'ampicilline (60 µg/ml; *Sigma*)
- Incuber overnight à 180 rpm et 37° C (*NBS, C24 Incubator Shaker*)
- Centrifuger à 3.500 rpm pendant 4' à 4° C (*Allegra X-22R, Beckman Coulter*) puis éliminer le surnageant
- Ajouter à chaque tube 300 µl de tampon P1 stocké à 4° C et mélanger par inversion
- Incuber pendant 10' à température ambiante
- Ajouter à chaque tube 300 µl de tampon P2 stocké à température ambiante et mélanger par inversion
- Incuber pendant 5' à température ambiante
- Ajouter à chaque tube 300 µl de tampon P3 stocké à 4° C et mélanger par inversion
- Incuber sur glace pendant 30'
- Centrifuger à 13.000 rpm pendant 10' à 4° C (*Biofuge A, Heraeus*) et récupérer le surnageant
- Ajouter à chaque tube 0,7 volume d'isopropanol stocké à -20° C et mélanger par inversion afin de précipiter l'ADN
- Incuber pendant 30' à température ambiante
- Centrifuger à 13.000 rpm pendant 15' à température ambiante (*Biofuge 13, Heraeus*) et éliminer le surnageant
- Ajouter à chaque tube 500 µl d'éthanol 70 % stocké à -20° C afin de laver l'ADN
- Centrifuger à 13 000 rpm pendant 5' à 4° C (*Biofuge A, Heraeus*) et éliminer le surnageant
- Sécher les culots au Speedvac pendant 15'
- Ajouter à chaque tube 25 µl de Tris 10 mM pH 8 afin de resuspendre l'ADN
- Placer sur agitateur rotatif overnight à 4° C
- Doser l'ADN par électrophorèse sur gel d'agarose 1 % avec du SL comme standard de taille

10. Séquençage

Matériels et solutions

- Mélange de séquençage (8 x 10 µl)
 - 32 µl CEQ DTCS Quick Start Kit (*Beckman Coulter* #608120)
 - 8 µl d'amorces T3 ou T7 non marquées 10 µM (*Sigma*)
 - 40 µl d'eau milliQ
- Solution stop (5 µl)
 - 2 µl acétate de sodium 3M pH 5,2 (*Sigma* #S7899)
 - 0,4 µl EDTA 0,5M pH 8 (*Sigma* #E7889)
 - 1µl glycogène (*Roche* #10901393001)
 - 1,6 µl d'eau milliQ
- Ethanol absolu stocké à -20° C (*Sigma*)
- Ethanol 70 % stocké à -20° C (*Sigma*)
- Sample Loading Solution (SLS, *Beckman Coulter* #608082)

Protocole

- Réaction de séquençage
 - Mélanger l'équivalent de 100 ng d'ADN avec de l'eau milliQ afin d'atteindre un volume final de 10 µl
 - Dénaturer le mélange pendant 15' à 95° C (thermocycler)
 - Transférer immédiatement sur glace pendant 3'
 - Ajouter à chaque échantillon 10 µl du mélange de séquençage
 - Lancer une réaction de séquençage suivant ces paramètres : 90'' à 92° C – [20'' à 96° C – 20'' à 50° C – 4' à 60° C] x 40 – 4° C à l'infini
- Précipitation de l'ADN (*voir section 8*)

La précipitation de l'ADN a été réalisée selon le protocole expliqué dans la section 8.

L'ADN a finalement été resuspendu dans 40 µl de SLS et analysé par un séquenceur à capillaires (CEQ 8000 fragment analyser, *Beckman Coulter*)

1. Design du pool d'ADNg

L'ADN de chicorée utilisé pour la recherche de polymorphisme est un superpool d'ADNg composé de 86 échantillons. Les feuilles au stade rosette de trois individus pour chacune des 29 variétés à la base de la sélection de *Cichorium intybus* L. ont été utilisées pour l'extraction d'ADN. Parmi ces 86 échantillons, 65 ont été extraits avec succès à l'URBV. Cependant, l'ADN apparaissait brun et était négatif à l'amplification, malgré des résultats positifs au dosage sur gel d'agarose et par spectrophotométrie 260/280 nm. Les échantillons ont donc été purifiés à l'aide du kit **NucleoSpin Extract II** (Macherey-Nagel #740 609.50), permettant finalement leur amplification. L'ADN des 21 échantillons manquants nous a été fourni par Warcoing S.A.. 3 µl de chacun des 86 échantillons ont été mélangés pour former un superpool d'ADN qui a été amplifié avec succès jusqu'à un ratio ADN/volume PCR de 1/1000, ce qui représente approximativement 50 pg d'ADN génomique par réaction.

2. Sélection des gènes candidats

À partir des informations fournies par la littérature, la sélection des gènes candidats a été réalisée sur base de la fonction des séquences impliquées dans la floraison d'*Arabidopsis thaliana*. 82 séquences différentes liées à la floraison ont ainsi été récupérées dans la base de données TAIR (<http://www.arabidopsis.org/>) en utilisant le mot-clé « flowering » et le numéro de référence des loci (ex : AT1G04400.1). Ces séquences ont ensuite été comparées à une base de données locale, créée à l'aide du logiciel **SeqTools**, qui contenait les séquences EST dérivées des 23.000 unigènes de *Cichorium intybus* L. Les séquences EST de chicorée montrant les meilleurs résultats de T-BlastX et les fonctions les plus intéressantes dans les voies de floraison ont alors été sélectionnées pour la conception d'amorces (**Tableau 1**).

3. Conception d'amorces

Les séquences ADN utilisées pour la conception des amorces ciblant les gènes candidats sélectionnés sont des séquences EST (Expressed Sequence Tag). Par définition, une séquence EST est une séquence complémentaire à l'ARNm du gène étudié. L'amplification des gènes candidats a quant à elle été réalisée sur un superpool d'ADNg, ce qui implique que la taille des produits amplifiés peut être différente de celle attendue, suite à la présence d'introns. En effet, la taille attendue variait entre 300-500 pb. Or, l'amplification de certains gènes candidats a donné des produits d'amplification de plus de 1000 pb. D'autres n'ont jamais été amplifiés, sans doute suite à une distance trop élevée entre les amorces de par la présence d'introns.

20 gènes candidats ont ainsi été sélectionnés dans un premier temps, pour lesquels 20 paires d'amorces ont été conçues. 17 paires d'amorces ont amplifié avec succès les gènes candidats, dont 5 ayant une séquence de plus de 1000 pb. Ces séquences de plus de 1000pb ne sont pas intéressantes à cloner car leur taille n'est pas compatible avec les analyses EcoTILLING ultérieures dont le standard de taille limite la détection des fragments à 1000 pb. C'est pourquoi 7 paires d'amorces supplémentaires ont ensuite été redéfinies, dont une ciblant un nouveau gène candidat. 4 de ces paires d'amorces ont amplifié avec succès les gènes candidats, dont 2 ayant une séquence de plus de 1000 pb. Donc, au final, 21 gènes candidats ont été amplifiés avec succès et les 14 paires d'amorces donnant des produits d'amplification inférieurs à 1000 pb ont été sélectionnées (**Tableau 2**).

4. Clonage des gènes candidats

Parmi les 21 gènes candidats amplifiés sur le superpool d'ADNg, 14 ont donc été clonés. Afin d'accélérer le clonage des gènes candidats, les étapes de restriction et de ligation des produits PCR ont été réalisées simultanément. Avant d'utiliser cette technique, différents protocoles de restriction et de ligation des produits PCR ont été réalisés sans succès. Malgré une analyse approfondie de la procédure et des réactifs utilisés, la raison pour laquelle cela a échoué systématiquement reste obscure. Le protocole de restriction-ligation simultanée pourrait sembler plus compliqué mais cette technique s'inspire de celle couramment utilisée en AFLP. Ce protocole de clonage orienté a ainsi été rendu possible par l'utilisation de sites de restriction compatibles dans le vecteur et dans l'insert. Ainsi, les sites de restriction XhoI (CTCGAG) et XbaI (TCTAGA) dans le vecteur sont compatibles avec les sites de restriction SalI (GTCGAC) et SpeI (ACTAGT) aux extrémités de l'insert, respectivement. En mettant l'insert en présence des enzymes de restriction SalI et SpeI, d'une ligase et du vecteur préalablement digéré par les enzymes de restriction XhoI et XbaI, il est ainsi possible de réaliser toutes les étapes en une seule fois et dans un seul tube. Ainsi, grâce aux adaptations apportées à cette technique, celle-ci est devenue totalement fonctionnelle et a évidemment permis d'accélérer le clonage des gènes candidats.

La transformation d'*E. coli* TOP 10F' thermocompétentes avec les produits de restriction-ligation a ensuite été réalisée par choc thermique. Les 14 gènes candidats clonés ont alors été criblés par PCR sur 64 colonies bactériennes. Le criblage par PCR a été effectué à l'aide des amorces marquées T3 (D4) et T7 (D3), permettant ainsi de marquer tous les produits PCR pour les analyses EcoTILLING ultérieures, ce qui fait encore gagner du temps et de l'argent. Les séquences universelles T3 et T7 sont localisées de part et d'autre du site multiple de clonage (MCS) dans le plasmide pBluescript II KS+ : la séquence T3 est localisée à 46 nucléotides du site de restriction XhoI et la séquence T7 à 52 nucléotides du site de restriction XbaI (**Figure 20**). La disruption du gène de la β -galactosidase dans le vecteur suite à l'insertion du produit PCR permet de discriminer les colonies positives (blanches avec insert) et négatives (bleues sans insert). Le nombre de colonies positives était habituellement proche de 100%, excepté pour le gène candidat AT5G57380.1 (72%) et AT6G65070 (89%) (**Tableau 3**).

5. Analyses EcoTILLING

Les produits PCR marqués issus du criblage des colonies positives ont ensuite été utilisés pour les analyses EcoTILLING qui fournissent des informations préliminaires concernant la présence de SNP dans le fragment analysé, permettant ainsi de sélectionner différents clones pour le séquençage.

5.1. Analyses EcoTILLING en pools

Les premières analyses EcoTILLING ont été effectuées sur les produits PCR poolés par colonne et par ligne (pools de 8 échantillons). Le principe de l'EcoTILLING est de comparer chaque échantillon à une référence. Le fait de pooler les échantillons doit donc permettre d'accélérer les analyses en localisant rapidement la séquence présentant les SNP d'intérêt.

Six gènes candidats ont ainsi été analysés en pools par EcoTILLING. L'amplification des ces gènes candidats a été réalisée, lors du criblage, à l'aide de l'amorce T3 marquée et de l'amorce T7 non marquée. À l'origine, l'EcoTILLING est réalisé à l'aide de deux amorces marquées, ce qui permet normalement de mettre en évidence une complémentarité entre les fragments détectés. Dans notre cas, la majorité des analyses a été effectuée à l'aide d'une seule amorce marquée car l'utilisation simultanée des fluorochromes sélectionnés engendrait pas mal de bruit de fond à l'analyse. Ainsi, seul le produit PCR du gène candidat AT1G79730.1 a été marqué aux deux extrémités.

Sur base des résultats obtenus, huit clones bactériens de chaque gène candidat (douze dans le cas du gène candidat AT1G79730.1) ont été sélectionnés pour le séquençage. Cette sélection s'est basée sur le nombre élevé de SNP ou la présence de SNP originaux détectés dans la séquence du clone bactérien (**Tableaux 4 à 9**).

SNP (pb)	Clones							
	4	7	10	22	30	33	34	38
72	72	72	72	72	72	72	72	72
94	94	94	94		94		94	94
98				98				
110						110		
140					140			
146					146			
153								153
164			164				164	
185	185			185	185		185	185
194					194		194	194
198				198		198	198	198
245			245					
261			261					
289	289							
294							294	
309				309				
324	324							
329	329		329					
340								340
377						377		
524			524					

L'amplification du gène candidat AT5G15850 (CO-like) donne un produit PCR de 605 pb. 39 SNP différents ont été détectés dans l'ensemble du pool. Huit clones bactériens différents, qui présentent au total 21 SNP distincts, ont été sélectionnés pour le séquençage.

Tableau 4 : SNP mis en évidence dans les huit clones bactériens sélectionnés pour le gène candidat AT5G15850 (CO-like)

SNP (pb)	Clones							
	1	10	24	29	35	39	46	49
63							63	
73	73	73	73	73	73	73		73
78							78	
85	85	85		85	85	85	85	85
90				90				
110	110							
113	113			113	113	113	113	113
116				116	116	116		116
120	120				120	120		120
143						143		
165	165		165	165	165	165	165	165
228							228	
276						276		
301	301	301		301	301	301		301
311	311		311	311	311	311	311	311
379			379	379	379		379	379
393	393	393			393			393
405			405					
415	415	415	415	415	415		415	415
429	429		429	429	429		429	429
436					436			
457	457							
472		472						
476					476			
492			492					
511								511
519				519				

L'amplification du gène candidat AT1G72050 (ELF6) donne un produit PCR de 632 pb. 36 SNP différents ont été détectés dans l'ensemble du pool. Huit clones bactériens différents, qui présentent au total 27 SNP distincts, ont été sélectionnés pour le séquençage.

Tableau 5 : SNP mis en évidence dans les huit clones bactériens sélectionnés pour le gène candidat AT1G72050 (ELF6)

SNP (pb)	Clones							
	6	13	22	26	32	36	49	64
63							63	
86	86	86	86		86	86	86	86
92				92				
102				102				
112						112		
122	122		122	122	122			
125			125					
134		134				134		
151					151			
189						189		
227						227		
253			253					
260						260		
293								293
299								299
310		310						
380	380							
408							408	

L'amplification du gène candidat AT3G24440 (VIL1) donne un produit PCR de 532 pb. 26 SNP différents ont été détectés dans l'ensemble du pool. Huit clones bactériens différents, qui présentent au total 18 SNP distincts, ont été sélectionnés pour le séquençage.

Tableau 6 : SNP mis en évidence dans les huit clones bactériens sélectionnés pour le gène candidat AT3G24440 (VIL1)

SNP (pb)	Clones											
	3	20	22	26	28	29	32	46	48	51	52	54
65							65					
105		105										
112		112	112	112	112	112	112		112			
129	129		129	129		129	129	129	129			
146	146											
151											151	
186								186				186
198										198		
200	200	200	200					200		200	200	200
235					235							
238			238									
240			240									
248	248		248					248	248	248		248
262			262									
297		297	297	297	297			297			297	297
323						323						
356				356	356	356						
378				378	378	378				378		
394												394
423									423			
462				462								

L'amplification du gène candidat AT1G79730.1 (ELF7) donne un produit PCR de 556 pb. Douze clones bactériens différents présentant les 21 SNP distincts détectés dans l'ensemble du pool ont finalement été sélectionnés pour le séquençage.

Tableau 7 : SNP mis en évidence dans les douze clones bactériens sélectionnés pour le gène candidat AT1G79730.1 (ELF7)

SNP (pb)	Clones							
	19	37	44	52	53	54	60	62
63		63		63	63	63	63	63
70		70			70	70		70
86								86
89				89				
99					99			
103						103		
105		105					105	
108					108	108		
114				114			114	
118							118	
120		120						
128								128
136					136			
147					147			
150					150			
163		163		163	163	163		
175	175	175		175	175	175	175	175
185	185							
196					196	196		196
213					213	213		213
218								218
271				271		271		
389	389							

L'amplification du gène candidat AT5G57380.1 (VIN3) donne un produit PCR de 492 pb. 27 SNP différents ont été détectés dans l'ensemble du pool. Huit clones bactériens différents, qui présentent au total 23 SNP distincts, ont été sélectionnés pour le séquençage.

Tableau 8 : SNP mis en évidence dans les huit clones bactériens sélectionnés pour le gène candidat AT5G57380.1 (VIN3)

		Clones							
		4	18	20	22	29	36	46	48
SNP (pb)	61		61	61	61			61	
	67		67	67	67			67	67
	74	74				74	74	74	74
	77		77	77					
	82								82
	87					87	87	87	87
	90			90	90				
	105			105					
	108	108							
	114					114	114	114	114
	118		118	118	118	118	118	118	118
	120			120	120		120	120	
	130		130	130	130	130		130	130
	175				175				
	227				227			227	
	243				243				
	258				258				
	272				272	272		272	272
	279				279			279	
	334		334	334	334	334	334	334	334
	368							368	
	421				421				
	486		486						
	490					490			
	581			581					
	774			774					
	811					811			
	841	841							

L'amplification du gène candidat AT6G65070 (FLC) donne un produit PCR de 905 pb. 40 SNP différents ont été détectés dans l'ensemble du pool. Huit clones bactériens différents, qui présentent au total 28 SNP distincts, ont été sélectionnés pour le séquençage.

Tableau 9 : SNP mis en évidence dans les huit clones bactériens sélectionnés pour le gène candidat AT6G65070 (FLC)

5.2. Analyses EcoTILLING en « un contre un »

Les analyses EcoTILLING suivantes ont été réalisées sur les clones traités individuellement en les comparant à une référence. Le fait d'avoir poolé nos échantillons, dans un premier temps, devait accélérer les analyses, mais ce procédé se révéla par la suite moins avantageux que prévu. En effet, ce système induit pas mal de bruit de fond et de faux positifs et ne permet de mettre en évidence avec certitude que les clones contenant du polymorphisme rare. Ces SNP ont alors une probabilité moindre de se retrouver chez les individus actuels et donc de pouvoir être potentiellement utilisés dans la sélection de ces individus. C'est pourquoi nous avons finalement changé d'optique. Ainsi, en analysant chacun des 64 échantillons issus du screening de chaque gène candidat contre une référence prise au hasard, le bruit de fond devait diminuer et le polymorphisme non rare être révélé. Les SNP présentant alors les plus hautes fréquences au sein des individus à la base de la sélection de la chicorée (superpool d'ADNg) pour la séquence étudiée pourraient être analysés par la suite chez les individus actuels afin de vérifier leur conservation ou leur perte au cours de la sélection.

Douze gènes candidats ont ainsi été analysés en « un contre un » par EcoTILLING. À l'instar des analyses précédentes, l'amplification des ces gènes candidats a été réalisée, lors du screening, à l'aide de l'amorce T3 marquée et de l'amorce T7 non marquée. Seul le produit PCR du gène candidat AT2G29950 a été marqué aux deux extrémités (**Figure 21**). Les produits PCR des 64 clones bactériens issus du criblage de chaque gène candidat ont donc été comparés à une référence choisie arbitrairement. Il s'agissait du produit PCR du clone 5 dans le cas des gènes candidats AT2G29950, AT5G15840.1, AT6G65070 et du clone 16 dans le cas des autres gènes candidats.

Sur base des résultats obtenus, quatre à sept clones bactériens de chaque gène candidat ont été sélectionnés pour le séquençage. Cette sélection s'est basée sur la fréquence relative des SNP détectés dans la séquence du clone bactérien. Cette fréquence reste relative car la totalité des 64 échantillons analysés pour chaque gène candidat en EcoTILLING ne fournit pas des sorties exploitables à l'analyse, le problème provenant généralement d'un manque d'échantillon ou de standard de taille suite aux manipulations. Les clones bactériens ont donc été sélectionnés pour le séquençage sur base des SNP les plus fréquemment observés (> 6%) (**Tableaux 10 à 18**). Les clones sélectionnés sont parfois redondants d'un point de vue de leur haplotype, ceci afin d'obtenir une confirmation des SNP observés par séquençage en comparant plusieurs séquences. A l'inverse, le clone servant de référence ne présentait évidemment aucun SNP par rapport à lui-même, mais il a été automatiquement sélectionné pour le séquençage.

Gène candidat	Produit PCR	SNP	Fréquence	SNP / Clones	05vs05	17vs05	21vs05	22vs05	41vs05
AT2G29950	590 pb	75	6,9%	75			75		
		135	25,9%	135				135	135
		163	8,6%	163			163		163
		165	8,6%	165		165			
		249	39,7%	249		249		249	249

Tableau 10 : SNP mis en évidence dans les cinq clones bactériens sélectionnés pour le gène candidat AT2G29950 (ELF4 – 590 pb). Ces SNP ont été sélectionnés parmi les 22 SNP différents détectés dans les 58 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence	SNP / Clones	16vs16	25vs16	28vs16	39vs16	44vs16	45vs16
AT3G10390.1	533 pb	114	90,6%	114		114	114		114	114
		115	50,9%	115			115			115
		181	94,3%	181		181	181		181	181
		419	73,6%	419		419	419			419
		485	18,9%	485		485	485			485

Tableau 11 : SNP mis en évidence dans les six clones bactériens sélectionnés pour le gène candidat AT3G10390.1 (FLD – 533 pb). Ces SNP ont été sélectionnés parmi les 9 SNP différents détectés dans les 53 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence
AT1G22770.1	484 pb	72	33,9%
		170	46,8%
		171	43,5%
		173	33,9%
		177	30,6%
		247	17,7%
		266	30,6%
		330	33,9%
		435	35,5%

SNP / Clones	04vs16	07vs16	10vs16	11vs16	13vs16	16vs16	52vs16
72					72		72
170	170	170		170	170		170
171	171	171		171			
173	173	173		173			
177					177		177
247	247			247			
266					266		266
330	330	330		330			
435		435	435				

Tableau 12 : SNP mis en évidence dans les sept clones bactériens sélectionnés pour le gène candidat AT1G22770.1 (GI – 484 pb). Ces SNP ont été sélectionnés parmi les 28 SNP différents détectés dans les 62 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence
AT5G15840.1	622 pb	286	60,3%
		316	14,3%
		350	27%
		509	81%

SNP / Clones	05vs05	06vs05	13vs05	14vs05	22vs05	64vs05
286		286		286	286	
316			316	316		
350		350			350	
509		509	509	509	509	

Tableau 13 : SNP mis en évidence dans les six clones bactériens sélectionnés pour le gène candidat AT5G15840.1 (CO – 622 pb). Ces SNP ont été sélectionnés parmi les 14 SNP différents détectés dans les 63 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence
AT5G65070	903 pb	75	12,1%
		115	48,3%
		130	46,6%
		273	70,7%
		561	6,9%

SNP / Clones	05vs05	07vs05	16vs05	52vs05	53vs05
75					75
115		115		115	
130		130	130	130	
273		273		273	
561		561			

Tableau 14 : SNP mis en évidence dans les quatre clones bactériens sélectionnés pour le gène candidat AT5G65070 (FLC – 903 pb). Ces SNP ont été sélectionnés parmi les 16 SNP différents détectés dans les 58 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence
AT1G04400.1	444 pb	115	50%
		123	7,7%
		128	7,7%
		229	32,7%
		280	15,4%
		380	36,5%
		440	46,2%

SNP / Clones	16vs16	21vs16	35vs16	37vs16	43vs16	47vs16
115		115		115		
123			123	123		
128					128	128
229		229			229	229
280			280	280		
380		380	380	380		
440		440	440	440	440	

Tableau 15 : SNP mis en évidence dans les six clones bactériens sélectionnés pour le gène candidat AT1G04400.1 (CRY2 – 444 pb). Ces SNP ont été sélectionnés parmi les 12 SNP différents détectés dans les 52 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence
AT1G25540.1	705 pb	69	7,8%
		164	17,6%
		305	7,8%
		451	23,5%
		650	19,6%

SNP / Clones	08vs16	16vs16	21vs16	26vs16	30vs16	54vs16	58vs16
69	69			69			
164			164		164		
305			305		305		
451			451			451	451
650						650	650

Tableau 16 : SNP mis en évidence dans les sept clones bactériens sélectionnés pour le gène candidat AT1G25540.1 (PFT1 – 705 pb). Ces SNP ont été sélectionnés parmi les 36 SNP différents détectés dans les 51 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence
AT5G24470.1	943 pb	95	13,7%
		612	52,9%
		705	41,2%
		734	23,5%
		900	27,5%

SNP / Clones	03vs16	06vs16	16vs16	22vs16
95	95	95		
612	612	612		612
705	705	705		705
734				734
900		900		

Tableau 17 : SNP mis en évidence dans les quatre clones bactériens sélectionnés pour le gène candidat AT5G24470.1 (APRR5 – 943 pb). Ces SNP ont été sélectionnés parmi les 58 SNP différents détectés dans les 51 produits PCR analysés.

Gène candidat	Produit PCR	SNP	Fréquence
AT4G16280.1	530 pb	149	50,8%
		342	23,7%
		435	44,1%
		440	39%
		468	37,3%
		474	54,2%
		477	54,2%
		523	35,6%
		528	22%

SNP / Clones	03vs16	05vs16	07vs16	08vs16	10vs16	16vs16	21vs16
149		149	149		149		149
342		342	342				342
435		435	435		435		435
440		440	440		440		440
468	468			468			468
474	474		474	474			474
477	477		477	477			477
523	523			523			
528			528				528

Tableau 18 : SNP mis en évidence dans les sept clones bactériens sélectionnés pour le gène candidat AT4G16280.1 (FCA – 530 pb). Ces SNP ont été sélectionnés parmi les 51 SNP différents détectés dans les 59 produits PCR analysés.

6. Analyses de séquençage

6.1. Première analyse

La première analyse a été effectuée sur les clones bactériens sélectionnés sur base des analyses EcoTILLING en pools. Pour rappel, cette sélection avait pour but de mettre en évidence un nombre élevé de SNP ou la présence de SNP originaux. Les clones bactériens ont été extraits par miniprep et séquencés au laboratoire. La quantité importante de polymorphisme observé au sein de ces séquences a parfois rendu difficile l'interprétation des résultats.

➤ Gène candidat AT5G15850 (COL)

Le gène candidat AT5G15850 correspond au gène CO-like qui encode une protéine nucléaire contrôlant l'expression du gène FT dans la voie dépendante de la photopériode. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (500 pb) et la séquence consensus d'ADNg de chicorée (500 pb) montre une bonne identité de séquence (99,4 %). Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-53} avec une protéine CO-like chez *Solanum lycopersicum*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 5 séquences différentes ont pu être obtenues, parmi les 8 clones bactériens séquencés, avec en moyenne un SNP toutes les 125 pb (**Tableau 19 (A)**).

Le produit PCR est localisé dans un seul exon présentant 4 SNP (**Figure 22 (A)**). Un de ces SNP a induit une modification dans la séquence protéique (I → V) qui implique deux acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale (non polaire). Cette modification n'est pas localisée dans le domaine B-box de liaison à l'ADN de la protéine (**Figure 22 (B)**). Les quatre SNP détectés dans la séquence génomique correspondent aux pics EcoTILLING localisés aux positions 110, 140, 164 et 245 (**Tableau 19 (B)**).

➤ Gène candidat AT1G72050 (ELF6)

Le gène candidat AT1G72050 correspond au gène ELF6 dans lequel des mutations provoquent une floraison précoce chez *A. thaliana* pendant les jours longs. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (303 pb) et la séquence consensus d'ADNg de chicorée (522 pb) montre une identité de séquence de 99 % entre les régions codantes. Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-64} avec une protéine en doigt de zinc de type C2H2 qui est une protéine similaire à REF6 (RELATIVE OF EARLY FLOWERING 6) chez *A. thaliana*. L'ADNc du gène candidat encode donc bien une protéine apparentée à celle ciblée. Pour ce gène candidat, 7 séquences différentes ont pu être obtenues, parmi les 8 clones bactériens séquencés, avec en moyenne un SNP toutes les 21 pb (**Tableau 20 (A)**).

La séquence d'ADNg est composée de trois exons et deux introns (219 pb – **Figure 23 (A)**). Il y a neuf SNP détectés dans l'intron 1 et cinq SNP dans l'intron 2. Il y a six SNP détectés dans l'exon 2, dont cinq qui ont induit une modification dans la séquence protéique. Trois de ces modifications (N → S, V → A, A → V) impliquent des acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale, tandis que les deux autres modifications (P → T et H → N) impliquent des acides aminés présentant des propriétés différentes. Ces modifications ne sont pas localisées dans le domaine en doigt de zinc de liaison à l'ADN de la protéine (**Figure 23 (B)**). Il y a cinq SNP détectés dans l'exon 3, dont un qui a induit une modification dans la séquence protéique (H → N) impliquant deux acides aminés présentant des propriétés différentes. Cette modification est localisée dans le domaine de liaison à l'ADN, ce qui pourrait avoir un impact significatif sur le fonctionnement de la protéine. Enfin, trois codons stop ont été observés au sein de l'exon 2, ce qui laisse supposer la détection d'un pseudogène. Par définition, un pseudogène est un gène inactif au sein du génome qui présente une similarité de séquence avec un autre gène qui lui est actif.

La détection de pseudogène n'est donc pas particulièrement intéressante dans le contexte de notre recherche de gènes candidats fonctionnels. Parmi les vingt-cinq SNP détectés dans la séquence génomique, seize correspondent aux pics observés lors des analyses EcoTILLING en pools (**Tableau 20 (B)**).

➤ Gène candidat AT3G24440 (VIL1)

Le gène candidat AT3G24440 correspond au gène VIN3-like 1 (VIL1) qui encode un composant du complexe de remodelage de la chromatine impliqué dans la régulation du gène FLC. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (425 pb) et la séquence consensus d'ADNg de chicorée (422 pb) montre une bonne identité de séquence (97,6 %). Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-37} avec le gène VIN3 chez *A. thaliana*. L'ADNc du gène candidat encode donc bien une protéine dont la fonction est relativement proche de celle ciblée. Pour ce gène candidat, 6 séquences différentes ont pu être obtenues, parmi les 8 clones bactériens séquencés, avec en moyenne un SNP toutes les 35 pb (**Tableau 21 (A)**).

Le produit PCR est localisé dans un seul exon présentant 12 SNP (**Figure 24 (A)**). Trois de ces SNP ont induit une modification dans la séquence protéique, dont deux (L → S et D → G) impliquant des acides aminés présentant des propriétés différentes au niveau de leur chaîne latérale et une (D → E) des acides aminés présentant des propriétés similaires (chaîne latérale ionisée). Deux délétions nucléotidiques ont également induit l'apparition d'un codon stop, ce qui a généralement pour conséquence la production d'une protéine tronquée. Parmi les douze SNP détectés dans la séquence génomique, cinq correspondent aux pics EcoTILLING localisés aux positions 92, 102, 112, 125 et 253 (**Tableau 21 (B)**).

➤ Gène candidat AT1G79730.1 (ELF7)

Le gène candidat AT1G79730.1 correspond au gène ELF7 dans lequel des mutations provoquent une floraison précoce chez *Arabidopsis* pendant les jours longs. Deux populations de séquences similaires mais différentes ont été obtenues après séquençage, ce qui laisse supposer la détection de deux populations de séquences paralogues (**Figure 25**). Le résultat du BlastX entre ces séquences d'ADNg de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-14} avec le gène ATGRP7 chez *A. thaliana* qui encode une protéine de liaison à l'ARN riche en glycine impliquée dans le rythme circadien et la réponse au froid. La fonction de ce gène ne correspond donc pas à celle du gène ELF7 ciblé mais semble néanmoins intéressante dans l'étude des gènes régulant les voies de floraison. Les amorces supposées cibler le gène ELF7 chez la chicorée ont en fait été conçues sur une séquence consensus, ce qui peut expliquer l'amplification d'une séquence seulement apparentée à la séquence ciblée. En effet, la séquence consensus utilisée a été sélectionnée dans un contig formé par différentes séquences de chicorée obtenues après un T-BlastX comparant des séquences nucléotidiques d'*A. thaliana* et une base de données locale des 23.000 unigènes (EST) de *Cichorium intybus* L. Le meilleur résultat de Blast doit normalement être sélectionné pour concevoir les amorces mais, dans ce cas, cette séquence n'était pas dans le contig.

La première séquence paralogue donne une séquence consensus d'ADNg de chicorée de 441 pb. L'alignement de cette séquence avec la séquence d'ADNc de chicorée (328 pb) montre une identité de séquence de 81,1 % entre les régions codantes. La densité de SNP relative dans ces trois séquences paralogues est d'un SNP toutes les 73 pb (**Tableau 22 (A)**). La séquence d'ADNg est composée de deux exons et d'un intron (113 pb – **Figure 26 (A)**). Il y a six SNP détectés dans ces séquences paralogues, dont quatre qui sont localisés dans l'intron 1. Il y a un SNP détecté dans l'exon 1 et un SNP dans l'exon 2 qui a induit une modification dans la séquence protéique (R → Q) impliquant deux acides aminés présentant des propriétés différentes au niveau de leur chaîne latérale. Enfin, deux codons stop ont été observés au sein de la séquence, ce qui laisse supposer la détection d'un pseudogène (**Tableau 22 (B)**).

La deuxième séquence paralogue donna une séquence consensus d'ADNg de chicorée de 454 pb. L'alignement de cette séquence avec la séquence d'ADNc de chicorée (328 pb) montra une identité de séquence de 98,2 % entre les régions codantes. La densité de SNP relative dans ces quatre séquences paralogues est d'un SNP toutes les 91 pb (**Tableau 22 (A)**). La séquence d'ADNg est composée de deux exons et d'un intron (126 pb – **Figure 26 (A)**). Il y a cinq SNP détectés dans ces séquences paralogues, dont quatre qui sont localisés dans l'intron 1 et qui pourraient avoir un impact important sur la régulation de la protéine. Le seul SNP détecté dans l'exon 2 n'a induit aucune modification dans le domaine conservé de la protéine (**Figure 26 (B)**).

Aucun des différents SNP détectés dans la séquence génomique de ces deux populations paralogues ne correspond aux pics observés lors des analyses EcoTILLING en pools (**Tableau 22 (B)**).

➤ Gène candidat AT5G57380.1 (VIN3)

Le gène candidat AT5G57380.1 correspond au gène VIN3 qui encode un composant du complexe de remodelage de la chromatine impliqué dans la régulation du gène FLC. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (378 pb) et la séquence consensus d'ADNg de chicorée (378 pb) montre une bonne identité de séquence (98,4 %). Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-37} avec le gène VIN3 chez *A. thaliana*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, seulement 2 séquences différentes ont pu être obtenues, parmi les 8 clones bactériens séquencés, avec en moyenne un SNP toutes les 126 pb (**Tableau 23 (A)**).

Le produit PCR est localisé dans un seul exon présentant 3 SNP (**Figure 27 (A)**). Deux de ces SNP ont induit une modification dans la séquence protéique. Une de ces modifications (D ↔ N) implique deux acides aminés présentant des propriétés différentes au niveau de leur chaîne latérale, tandis que l'autre modification (V ↔ L) implique deux acides aminés présentant des propriétés similaires (non polaire). Parmi les trois SNP détectés dans la séquence génomique, un seul correspond au pic EcoTILLING localisé à la position 150 (**Tableau 23 (B)**).

➤ Gène candidat AT5G65070 (FLC)

Le gène candidat AT5G65070 correspond au gène FLC qui encode un facteur de transcription « MADS box » inhibant la floraison. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (368 pb) et la séquence consensus d'ADNg de chicorée (782 pb) montre une identité de séquence de 91,3 % entre les régions codantes. Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-30} avec un gène FLC-like chez *Vitis vinifera*. L'ADNc du gène candidat encode donc bien une protéine dont la fonction est relativement proche de celle ciblée. Pour ce gène candidat, 7 séquences différentes ont pu être obtenues, parmi les 8 clones bactériens séquencés, avec en moyenne un SNP toutes les 37 pb (**Tableau 24 (A)**).

La séquence d'ADNg est composée de six exons et cinq introns (414 pb – **Figure 28 (A)**). Il y a un SNP détecté dans l'intron 1, une délétion et quatre SNP dans l'intron 3 et quatre SNP dans l'intron 4. Il y a un SNP détecté dans l'exon 1 qui n'a induit aucune modification dans la séquence protéique. Il y a un SNP dans l'exon 2 qui a induit une modification dans la séquence protéique impliquant deux acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale (chaîne latérale ionisée). Il y a deux délétions nucléotidiques détectées dans l'exon 3 qui ont induit l'apparition d'un codon stop, ce qui a généralement pour conséquence la production d'une protéine tronquée. Il y a trois SNP détectés dans l'exon 4, dont deux qui ont induit une modification dans la séquence protéique. Une de ces modifications (L → I) implique deux acides aminés présentant des propriétés similaires (non polaire), tandis que l'autre modification a induit l'apparition d'un codon stop. Il y a quatre SNP détectés dans l'exon 6, dont trois qui ont induit une modification dans la séquence protéique. Deux de ces modifications (L → F et L → M) impliquent des acides aminés présentant des propriétés similaires, tandis que l'autre modification implique des acides aminés présentant des propriétés différentes. Enfin, un codon stop a été observé au sein de la séquence, ce qui laisse supposer la détection d'un pseudogène. Parmi les vingt-un SNP détectés dans la séquence génomique, sept correspondent aux pics EcoTILLING localisés aux positions 74, 105, 118, 130, 258, 272 et 368 (**Tableau 24 (B)**).

6.2. Deuxième analyse

La deuxième analyse a été effectuée sur les produits PCR correspondants aux clones bactériens sélectionnés sur base des analyses EcoTILLING en « un contre un ». Pour rappel, cette sélection avait pour but de mettre en évidence les SNP les plus fréquemment observés (>6 %) en se basant sur l'hypothèse que si leur fréquence « ancestrale » est suffisante, il est possible qu'ils soient toujours présents à l'heure actuelle. Certains des clones sélectionnés présentaient, lors des analyses EcoTILLING, des SNP supplémentaires à ceux dont la fréquence était supérieure à 6 %. Ces SNP ont parfois été confirmés lors du séquençage, mais pas dans tous les cas. Les produits PCR ont été purifiés et séquencés par **MACROGEN**. La totalité des séquences obtenues pour chacun des neuf gènes candidats analysés a pu être exploitée.

➤ Gène candidat AT2G29950 (ELF4)

Le gène candidat AT2G29950 correspond au gène ELF4 dans lequel des mutations provoquent une floraison précoce chez *Arabidopsis* pendant les jours longs. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (481 pb) et la séquence consensus d'ADNg de chicorée (481 pb) montre une bonne identité de séquence (99,6 %). Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-17} avec le gène ELF4 chez *A. thaliana*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 5 séquences différentes ont pu être obtenues, parmi les 5 clones bactériens séquencés, avec en moyenne un SNP toutes les 41 pb (**Tableau 25 (A)**).

Le produit PCR est localisé dans un seul exon présentant neuf SNP (**Figure 29 (A)**). Sept de ces SNP ont induit une modification dans la séquence protéique. Deux de ces modifications (K → E et F → L) impliquent des acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale. Ces modifications sont localisées dans le domaine conservé de la protéine et pourraient donc avoir un impact important sur son fonctionnement (**Figure 29 (B)**). Les cinq autres modifications impliquent des acides aminés présentant des propriétés différentes, dont deux (S → R et E → G) qui sont également localisées dans le domaine conservé de la protéine. Enfin, un codon stop a été observé au sein de la séquence, suggérant la détection d'un pseudogène potentiel. Parmi les neuf SNP détectés dans la séquence génomique, six correspondent aux pics EcoTILLING localisés aux positions 75, 135, 165, 249, 303 et 397 dont la fréquence était de 6,9 %, 25,9 %, 8,6 %, 39,7 %, 1,7 % et 1,7 % respectivement (**Tableau 25 (B)**).

➤ Gène candidat AT3G10390.1 (FLD)

Le gène candidat AT3G10390.1 correspond au gène FLD qui encode une protéine nucléaire impliquée dans le remodelage de la chromatine à la base de la régulation épigénétique du gène FLC. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (424 pb) et la séquence consensus d'ADNg de chicorée (424 pb) montre une bonne identité de séquence (98,6 %). Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-102} avec le gène FLD chez *A. thaliana*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 3 séquences différentes ont pu être obtenues, parmi les 6 clones bactériens séquencés, avec en moyenne un SNP toutes les 42 pb (**Tableau 26 (A)**).

Le produit PCR est localisé dans un seul exon présentant dix SNP (**Figure 30 (A)**). Trois de ces SNP ont induit une modification dans la séquence protéique impliquant des acides aminés présentant des propriétés différentes au niveau de leur chaîne latérale. Un quatrième SNP a induit l'apparition d'un codon stop, ce qui a généralement pour conséquence la production d'une protéine tronquée. Ces quatre modifications sont localisées dans le domaine fonctionnel de la protéine, ce qui pourrait avoir un impact significatif sur son fonctionnement (**Figure 30 (B)**). Parmi les dix SNP détectés dans la séquence génomique, trois correspondent aux pics EcoTILLING localisés aux positions 114, 181 et 419 dont la fréquence était de 90,6 %, 94,3 % et 73,6 % respectivement (**Tableau 26 (B)**).

➤ Gène candidat AT1G22770.1 (GI)

Le gène candidat AT1G22770.1 correspond au gène GI qui encode une protéine nucléaire contrôlant l'expression du gène intégrateur FT. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (373 pb) et la séquence consensus d'ADNg de chicorée (373 pb) montre une bonne identité de séquence (99,2 %). Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 100 % avec le gène GI chez *A. thaliana*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 7 séquences différentes ont pu être obtenues, parmi les 7 clones bactériens séquencés, avec en moyenne un SNP toutes les 22 pb (**Tableau 27 (A)**).

Le produit PCR est localisé dans un seul exon présentant dix-sept SNP (**Figure 31 (A)**). Cinq de ces SNP ont induit une modification dans la séquence protéique. Deux de ces modifications (K → R et F → L) impliquent des acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale, tandis que les trois autres modifications impliquent des acides aminés présentant des propriétés différentes. Parmi les dix-sept SNP détectés dans la séquence génomique, huit correspondent aux pics EcoTILLING localisés aux positions 72, 102, 170, 177, 187, 247, 266 et 330 dont la fréquence était de 33,9 %, 3 %, 46,8 %, 30,6 %, 3 %, 17,7 %, 30,6 % et 33,9 % respectivement (**Tableau 27 (B)**).

➤ Gène candidat AT5G15840.1(CO)

Le gène candidat AT5G15840.1 correspond au gène CO qui encode une protéine nucléaire avec deux doigts de zinc très similaires à la boîte-B contrôlant l'expression du gène intégrateur FT. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (411 pb) et la séquence consensus d'ADNg de chicorée (514 pb) montre une identité de séquence de 99 % entre les régions codantes. Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-91} avec un gène CO-like chez *Solanum tuberosum*, ainsi qu'une similarité de 10^{-72} avec une protéine en doigt de zinc de type boîte-B chez *A. thaliana*. L'ADNc du gène candidat encode donc bien une protéine dont la fonction est relativement proche de celle ciblée. Pour ce gène candidat, 6 séquences différentes ont pu être obtenues, parmi les 6 clones bactériens séquencés, avec en moyenne un SNP toutes les 47 pb (**Tableau 28 (A)**).

La séquence d'ADNg est composée de deux exons et un intron (103 pb – **Figure 32 (A)**). Il y a deux SNP détectés dans l'intron 1. Il y a deux SNP détectés dans l'exon 1 qui ont induit une modification dans la séquence protéique impliquant des acides aminés présentant des propriétés différentes au niveau de leur chaîne latérale. Il y a sept SNP détectés dans l'exon 2, dont quatre qui ont induit une modification dans la séquence protéique. Une de ces modifications (D → E), qui est observée deux fois au sein de l'exon, implique des acides aminés présentant des propriétés similaires (chaîne latérale ionisée), tandis que les deux autres modifications impliquent des acides aminés présentant des propriétés différentes. Parmi les onze SNP détectés dans la séquence génomique, six correspondent aux pics EcoTILLING localisés aux positions 169, 250, 286, 316, 502 et 509 dont la fréquence était de 1,5 %, 1,5 %, 60,3 %, 14,3 %, 1,5 % et 81 % respectivement (**Tableau 28 (B)**).

➤ Gène candidat AT5G65070 (FLC)

Le gène candidat AT5G65070 correspond au gène FLC qui encode un facteur de transcription « MADS box » inhibant la floraison. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (368 pb) et la séquence consensus d'ADNg de chicorée (781 pb) montre une identité de séquence de 91,3 % entre les régions codantes. Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-30} avec un gène FLC-like chez *Vitis vinifera*. L'ADNc du gène candidat encode donc bien une protéine dont la fonction est relativement proche de celle ciblée. Pour ce gène candidat, 5 séquences différentes ont pu être obtenues, parmi les 5 clones bactériens séquencés, avec en moyenne un SNP toutes les 49 pb (**Tableau 29 (A)**).

La séquence d'ADNg est composée de six exons et cinq introns (413 pb – **Figure 33 (A)**). Il y a une délétion dans l'intron 1, un SNP dans l'intron 2, trois SNP dans l'intron 4 et un SNP dans l'intron 5. Il y a trois SNP détectés dans l'exon 3, dont un qui a induit une modification dans la séquence protéique (D → E) impliquant deux acides aminés présentant des propriétés différentes au niveau de leur chaîne latérale. Il y a un SNP détecté dans l'exon 4 qui n'a induit aucune modification dans la séquence protéique. Il y a cinq SNP détectés dans l'exon 6, dont quatre qui ont induit une modification dans la séquence protéique. Une de ces modifications (F → T) implique deux acides aminés présentant des propriétés différentes, tandis que les trois autres modifications impliquent des acides aminés présentant des propriétés similaires. Enfin, un codon stop a été observé au sein de la séquence, suggérant la détection d'un pseudogène potentiel. Parmi les quinze SNP détectés dans la séquence génomique, cinq correspondent aux pics observés lors des analyses EcoTILLING en « un contre un » (**Tableau 29 (B)**).

Ce gène candidat a été analysé deux fois en EcoTILLING (en pools et en « un contre un »). Seuls trois SNP observés lors de la première analyse ont été confirmés au cours de la deuxième analyse. Il s'agit des SNP en position 22, 65 et 81 de la séquence génomique.

➤ Gène candidat AT1G04400.1 (CRY2)

Le gène candidat AT1G04400.1 correspond au gène CRY2 qui encode un des cryptochromes impliqué dans la perception de la lumière. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (325 pb) et la séquence consensus d'ADNg de chicorée (327 pb) montre une bonne identité de séquence (98,5 %). Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-111} avec le gène CRY2 chez *A. thaliana*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 4 séquences différentes ont pu être obtenues, parmi les 6 clones bactériens séquencés, avec en moyenne un SNP toutes les 47 pb (**Tableau 30 (A)**).

Le produit PCR est localisé dans un seul exon présentant sept SNP (**Figure 34 (A)**). Trois de ces SNP ont induit une modification dans la séquence protéique. Une de ces modifications (G → V) implique deux acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale (non polaire), tandis que les deux autres modifications impliquent des acides aminés présentant des propriétés différentes. Enfin, un codon stop a été observé au sein de la séquence, suggérant la détection d'un pseudogène potentiel. Parmi les sept SNP détectés dans la séquence génomique, quatre correspondent aux pics EcoTILLING localisés aux positions 115, 128, 229 et 280 dont la fréquence était de 50 %, 7,7 %, 32,7 % et 15,4 % respectivement (**Tableau 30 (B)**).

➤ Gène candidat AT1G25540.1 (PFT1)

Le gène candidat AT1G25540.1 correspond au gène PFT1 qui encode une protéine nucléaire impliquée dans la régulation du gène FT en fonction de la qualité de la lumière. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (371 pb) et la séquence consensus d'ADNg de chicorée (592 pb) montre une identité de séquence de 98,7 % entre les régions codantes. Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-44} avec le gène PFT1 chez *A. thaliana*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 3 séquences différentes ont pu être obtenues, parmi les 7 clones bactériens séquencés, avec en moyenne un SNP toutes les 296 pb (**Tableau 31 (A)**).

La séquence d'ADNg est composée de quatre exons et trois introns (221 pb – **Figure 35 (A)**). Il y a un SNP détecté dans l'intron 2 et un SNP dans l'exon 2 qui a induit une modification dans la séquence protéique impliquant deux acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale (polaire). Les deux SNP détectés dans la séquence génomique correspondent aux pics EcoTILLING localisés aux positions 305 et 451 dont la fréquence était de 7,8 % et 23,5 % respectivement (**Tableau 31 (B)**).

➤ Gène candidat AT5G24470.1 (APRR5)

Le gène candidat AT5G24470.1 correspond au gène APRR5 impliqué dans la régulation de l'horloge circadienne. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (414 pb) et la séquence consensus d'ADNg de chicorée (830 pb) montre une identité de séquence de 100 % entre les régions codantes. Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-98} avec le gène APRR5 chez *A. thaliana*. L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 4 séquences différentes ont pu être obtenues, parmi les 4 clones bactériens séquencés, avec en moyenne un SNP toutes les 119 pb (**Tableau 32 (A)**).

La séquence d'ADNg est composée de quatre exons et trois introns (416 pb – **Figure 36 (A)**). Il y a six SNP détectés dans l'intron 1 et un SNP dans l'intron 2. Aucun SNP n'a été détecté dans les exons, n'induisant donc aucune modification dans le domaine fonctionnel de la protéine (**Figure 36 (B)**). Parmi les sept SNP détectés dans la séquence génomique, quatre correspondent aux pics EcoTILLING localisés aux positions 401, 612, 705 et 734 dont la fréquence était de 4 %, 52,9 %, 41,2 % et 23,5 % respectivement (**Tableau 32 (B)**).

➤ Gène candidat AT4G16280.1 (FCA)

Le gène candidat AT4G16280.1 correspond au gène FCA impliqué dans la régulation négative du gène FLC. Le résultat de l'alignement entre la séquence d'ADNc de chicorée (264 pb) et la séquence consensus d'ADNg de chicorée (416 pb) montre une identité de séquence de 99,6 % entre les régions codantes. Le résultat du BlastX entre la séquence d'ADNc de chicorée et la base de données protéique de NCBI montre une similarité de 10^{-12} avec le gène FCA chez *Triticum aestivum* (blé). L'ADNc du gène candidat encode donc bien la protéine ciblée. Pour ce gène candidat, 4 séquences différentes ont pu être obtenues, parmi les 7 clones bactériens séquencés, avec en moyenne un SNP toutes les 83 pb (**Tableau 33 (A)**).

La séquence d'ADNg est composée de trois exons et deux introns (152 pb – **Figure 37 (A)**). Il y a un SNP détecté dans l'intron 1 et un SNP dans l'intron 2. Il y a une délétion nucléotidique détectée dans l'exon 1 qui a induit l'apparition d'un codon stop, ce qui a généralement pour conséquence la production d'une protéine tronquée. Il y a deux SNP dans l'exon 2, dont un qui a induit une modification dans la séquence protéique impliquant deux acides aminés présentant des propriétés similaires au niveau de leur chaîne latérale (non polaire). Ces différentes modifications ne sont pas localisées dans le domaine d'interaction protéique (**Figure 37 (B)**). Enfin, un codon stop a été observé au sein de la séquence, suggérant la détection d'un pseudogène potentiel. Parmi les cinq SNP détectés dans la séquence génomique, trois correspondent aux pics EcoTILLING localisés aux positions 149, 342 et 435 dont la fréquence était de 50,8 %, 23,7 % et 44,1 % respectivement (**Tableau 33 (B)**).

Les deux analyses de séquençage nous ont donc permis de mettre en évidence treize gènes de chicorée potentiellement orthologues à ceux d'*A. thaliana*. Les séquences consensus d'ADNg similaires mais différentes obtenues pour le gène candidat AT1G79730.1 semblent quant à elles paralogues. Les pseudogènes mis en évidence représentent sans doute la source d'informations la moins intéressante pour les analyses ultérieures. Une majorité de mutations silencieuses a également été détectée dans ces séquences. Celles-ci auront probablement moins d'impact sur le phénotype par rapport aux mutations faux-sens, d'autant plus si ces dernières se situent dans un domaine clé de la protéine comme observé pour les gènes candidats AT1G72050, AT2G29950 et AT3G10390.1. Les mutations non-sens ont également une plus grande probabilité d'influencer le phénotype, étant donné qu'elles sont généralement à l'origine d'une protéine tronquée. Le choix des gènes candidats les plus intéressants à analyser dans des études d'association ultérieures devrait donc se baser sur ces différents critères.

Au cours de ce mémoire, nous avons développé et testé une technique permettant de rechercher des marqueurs moléculaires ciblant les gènes de floraison chez *Cichorium intybus* L. Un grand nombre de gènes candidats potentiellement impliqués dans le contrôle de la floraison ont été identifiés sur base de la littérature. Les gènes orthologues potentiels de chicorée ont ensuite été récupérés dans une base de données contenant les séquences EST dérivées des 23.000 unigènes de *Cichorium intybus* L.

Vingt-et-une paires d'amorces ont été conçues pour la recherche de polymorphisme dans ces gènes candidats et testées avec succès sur une « core collection » d'ADNg issus des vingt-neuf variétés à la base de la sélection de la chicorée. L'intérêt d'une core collection est de disposer d'une représentation aussi large que possible de la diversité génétique présente au sein de la population étudiée, et ce à partir d'un nombre réduit de génotypes. Les produits d'amplification de quatorze des gènes candidats ont ensuite été clonés dans le plasmide pBluescript II KS+ grâce à une étape de restriction-ligation simultanée. L'utilisation de ce protocole inspiré et adapté de celui utilisé en AFLP nous a ainsi permis d'accélérer le clonage de nos gènes candidats. L'utilisation de bactéries *E. coli* TOP 10F' thermocompétentes pour la transformation par choc thermique nous a permis d'atteindre une efficacité de transformation proche de 100 %. Le criblage des colonies bactériennes ainsi obtenues a été réalisé à l'aide des amorces universelles T3/T7 marquées. L'utilisation de ces amorces nous a permis de marquer en une seule étape tous les produits PCR nécessaires aux analyses EcoTILLING ultérieures. Le protocole de clonage orienté des gènes candidats décrit ci-dessus peut donc être considéré comme rapide et totalement fonctionnel grâce aux différentes mises au point et adaptations apportées au cours du mémoire.

L'étape suivante a été la recherche du polymorphisme génétique de type SNP au sein des gènes candidats. Dans un premier temps, la détection des SNP a été réalisée par des analyses EcoTILLING sur des pools d'échantillons, et ce afin d'accélérer encore une fois les analyses. Le nombre de SNP détectés dans les séquences analysées en pools était particulièrement important, ce qui peut s'expliquer de différentes façons. Tout d'abord, le fait d'analyser une core collection d'individus anciens implique au départ la présence d'une grande diversité génétique. Ensuite, le fait d'avoir potentiellement des gènes paralogues dans ces analyses augmente encore la densité de SNP observés. En effet, le polymorphisme des gènes candidats doit nous permettre de discriminer les gènes orthologues, c-à-d des gènes présentant un même locus entre différentes espèces, et les gènes paralogues, c-à-d des gènes présentant des séquences similaires mais un locus différent au sein de la même espèce. Par définition, la présence de gènes paralogues dans l'analyse EcoTILLING va augmenter le bruit de fond suite à une amplification de gènes non spécifique. Enfin, le fait d'avoir poolé nos échantillons devait accélérer les analyses, mais ce système montra par la suite un inconvénient majeur qui est l'identification de clones présentant essentiellement du polymorphisme rare. Cela s'explique par le fait que si un SNP est présent dans chacun des échantillons poolés par colonne et par ligne, il passera inaperçu et ne sera donc pas mis en évidence, malgré sa fréquence élevée dans les séquences analysées. On suppose que les SNP plus rares ont une probabilité moindre de se retrouver chez les individus issus des variétés actuelles et qu'ils ne représentent donc pas la source de polymorphisme la meilleure à exploiter dans la sélection de ces individus. C'est pourquoi la suite des analyses EcoTILLING a été réalisée sur des échantillons individuels, afin de diminuer le bruit de fond et de révéler le polymorphisme plus fréquent.

Néanmoins, le nombre de SNP détectés dans les séquences analysées individuellement resta élevé, sans doute suite à l'énorme diversité génétique présente au sein de la core collection. C'est pourquoi les résultats obtenus sur celle-ci devront être par la suite confirmés par des analyses sur les individus issus de variétés actuelles, afin de vérifier si les SNP les plus fréquemment observés au sein de la core collection ont été conservés au cours de la sélection.

La dernière étape a été le génotypage des SNP détectés en EcoTILLING. Le séquençage des clones bactériens sélectionnés sur base des analyses EcoTILLING en pools a été effectué au laboratoire. Le séquençage des clones bactériens sélectionnés sur base des analyses EcoTILLING en « un contre un » a quant à lui été effectué par la firme coréenne MACROGEN. Ce séquençage a été réalisé directement à partir des produits PCR d'intérêt purifiés par leur soin. Au final, le séquençage de MACROGEN s'avéra donc plus efficace et moins coûteux que notre propre séquençage. Auparavant, deux autres techniques de purification des produits PCR ont été testées au laboratoire, afin de ne plus devoir extraire les clones bactériens par miniprep. Cela devait nous permettre d'accélérer la procédure et d'améliorer peut-être la qualité des résultats obtenus. Il s'agissait d'utiliser de l'Agarose Low Melting (Ma et al., 2008) ou un mélange d'exonucléase et de SAP (Werle et al., 1994) ; deux méthodes prometteuses qui ne furent finalement pas utilisées dans le cadre de nos analyses.

Seuls 40 % des SNP observés au séquençage avaient été détectés lors des analyses EcoTILLING, que ce soit en pools ou en « un contre un », ce qui montre bien qu'un certain bruit de fond persiste. La très grande diversité génétique observée au sein de la core collection représente donc une analyse préliminaire intéressante mais imprécise. En effet, ces analyses permettent d'avoir une idée du profil génétique de la population étudiée avant la sélection, ce qui représente des données intéressantes avant d'analyser le profil génétique des individus actuels après la sélection. Cependant, seules ces analyses ultérieures nous permettront de mettre en évidence des SNP potentiellement liés ou impliqués dans la variation du phénotype. Ainsi, en comparant le polymorphisme génétique d'individus actuels généalogiquement proches qui sont soit sensibles soit résistants à la montaison, il sera peut-être possible de mettre en évidence les SNP associés à cette variation phénotypique. En effet, en choisissant des individus qui sont génétiquement proches mais qui présentent des phénotypes très différents, les mutations détectées entre leurs génomes auront une plus grande probabilité d'être impliquées dans cette variation phénotypique, plutôt que dans la variabilité naturelle observée entre individus différents.

L'effet potentiel de ces SNP sur le phénotype doit donc encore être étudié. Dans un premier temps, il serait intéressant d'effectuer un alignement multiple des séquences obtenues avec d'autres séquences d'*Astéracées*. Cela pourrait permettre d'observer si certaines des mutations détectées ne se sont pas localisées dans des séquences conservées, et ce afin de cibler dans les analyses ultérieures les SNP ayant la plus grande probabilité d'avoir un impact sur le phénotype. L'impact de ces SNP et des changements d'acides aminés observés dans les séquences protéiques pourraient alors être approfondis, par la suite, par mutagenèse dirigée. En effet, l'observation de la structure (par cristallographie) et de la fonction (par des tests d'activité) de tels mutants permettraient de définir l'impact des SNP détectés sur la protéine produite. Enfin, l'impact du polymorphisme détecté dans les gènes candidats pourra être corrélé aux variations phénotypiques observées lors d'études d'association. Cela permettra de déterminer les SNP liés à la séquence responsable du phénotype ou, plus rarement, les SNP directement impliqués dans la variation phénotypique. Des expériences de phénotypage en champ seront finalement nécessaires à la validation de ces observations. Ces différentes étapes devraient alors permettre d'obtenir des marqueurs moléculaires exploitables en sélection assistée par marqueurs dans le domaine de la résistance à la montaison.

Dans cette optique, une technique d'analyse à haut débit dérivée de la méthode du Golden Gate Assay doit encore être mise au point au laboratoire.

Les différents résultats obtenus au cours de ce mémoire nécessitent donc des analyses supplémentaires plus approfondies pour pouvoir être exploités en sélection. Cependant, le protocole de clonage des gènes candidats et la technique de détection et de génotypage des SNP peuvent être considérés comme fonctionnels. Quelques mises au point dans le processus d'EcoTILLING devraient permettre de diminuer encore le bruit de fond, améliorant ainsi la sélection des individus intéressants à génotyper afin d'obtenir des résultats de séquençage les plus informatifs possibles. Ces différentes étapes pourraient donc être utilisées pour la recherche de polymorphisme dans de nombreux autres gènes candidats. En effet, il serait intéressant d'approfondir la recherche de gènes candidats potentiellement impliqués dans la régulation de la floraison. Cela pourrait se faire à partir de différentes données issues de la littérature ou des bases de données informatiques, telles que celles d'*Arabidopsis thaliana*, de *Cichorium intybus* L. ou des autres espèces proches de la chicorée (ex.: *Cichorium endivia*), d'*Helianthus* spp, de *Lactuca* spp. ou encore d'autres espèces d'Asteracées (ex. : pissenlit).

- BAE G., CHOI G., Decoding of light signals by plant phytochromes and their interacting proteins, Annual Review of Plant Biology, 59:281-311, 2008
- BAIS H.P., RAVISHANKAR G.A., Cichorium intybus L – cultivation, processing, utility, value addition and biotechnology, with an emphasis on current status and future prospects, Journal of the Science of Food and Agriculture, 81:467-484, 2001
- BOSS P.K., BASTOW R.M., MYLNE J.S., DEAN C., Multiple pathways in the decision to flower: Enabling, promoting, and resetting, The Plant Cell, 16:18-31, 2004
- CASAL J.J., Phytochromes, Cryptochromes, Phototropin: Photoreceptor Interactions in Plants, Photochemistry and Photobiology, 2000, 71:1-11
- CASTILLON A., SHEN H., HUQ E., Phytochrome Interacting Factors: central players in phytochrome-mediated light signaling networks, Trends in Plant Science, 12:514-521, 2007
- COMAI L., YOUNG K., TILL B.J., REYNOLDS S.H., GREENE E.A., CODOMO C.A., ENNS L.C., JOHNSON J.E., BURTNER C., ODDEN A.R., HENIKOFF S., Efficient discovery of DNA polymorphisms in natural populations by Ecotilling, The Plant Journal, 37:778-786, 2004
- CORBESIER L., VINCENT C., JANG S., FORNARA F., FAN Q., SEARLE I., GIAKOUNTIS A., FARRONA S., GISSOT L., TURNBULL C., COUPLAND G., FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis, Science, 316:1030-1033, 2007
- CORNELIS P., Guide de Bonnes Pratiques en Culture de Chicorée Industrielle, Centre Agricole Betterave-Chicorée (CABC), 2003
- CHRISTIE J.M., Phototropin blue-light receptors, Annual Review of Plant Biology, 58:21-45, 2007
- DEMEULEMEESTER M.A.C., DE PROFT M.P., In vivo and in vitro flowering response of chicory (Cichorium intybus L.): influence of plant age and vernalization, Plant Cell Reports, 18:781-785, 1999
- DE VIENNE D., Les marqueurs moléculaires en génétique et biotechnologies végétales, Editions INRA, Collections Mieux comprendre, 1998
- JACK T., Molecular and genetic mechanisms of floral control, The Plant Cell, Supplement: 1-17, 2004
- KWOK P-Y, CHEN X., Detection of Single Nucleotide Polymorphisms, Current Issues in Molecular Biology, 5:43-60, 2003
- LI Q-H, YANG H-Q, Cryptochrome signaling in plants, Photochemistry and Photobiology, 83:94 -101, 2007
- LIN C., Photoreceptors and Regulation of Flowering Time, Plant Physiology, 123:39-50, 2000
- LIN C., Blue light receptors and signal transduction, The Plant Cell, Supplement: 207-225, 2002
- LIN C., TODO T., The cryptochromes, Genome Biology, 6:220.1-220.9, 2005
- MA H., DIFAZIO S., An efficient method for purification of PCR products for sequencing, BioTechniques, 44:921-923, 2008
- MACKNIGHT R., DUROUX M., LAURIE R., DIJKWEL P., SIMPSON G., DEAN C., Functional significance of the alternative transcript processing of the Arabidopsis floral promoter FCA, The Plant Cell, 14:877-888, 2002
- MARQUARDT S., BOSS P.K., HADFIELD J., DEAN C., Additional targets of the Arabidopsis autonomous pathway members, FCA and FY, Journal of Experimental Botany, 57:3379-3386, 2006
- MOURADOV A., CREMER F., COUPLAND G., Control of flowering time: Interacting pathways as a basis for diversity, The Plant Cell, 14:111-130, 2002

- OLSZEWSKI N., SUN T-P, GUBLER F., Gibberellin signaling: biosynthesis, catabolism, and response pathways, The Plant Cell, Supplement: 61-80, 2002
- ÖRVAR B.L., SANGWAN V., OMANN F., DHINDSA R.S., Early steps in cold sensing by plant cells: the role of actin cytoskeleton and membrane fluidity, The Plant Journal, 23:785-794, 2000
- PENFIELD S., Temperature perception and signal transduction in plants, New Phytologist, 179:615-628, 2008
- PFLIEGER S., LEFEBVRE V., CAUSSE M., The candidate gene approach in plant genetics : a review, Molecular Breeding, 7:275-291, 2001
- PUTTERILL J., Flowering in time: genes controlling photoperiodic flowering in Arabidopsis, The Philosophical Transactions of the Royal Society, 356:1761-1767, 2001
- SCHMITZ R.J., AMASINO R.M., Vernalization: A model for investigating epigenetics and eukaryotic gene regulation in plants, Biochimica et Biophysica Acta, 1769:269-275, 2007
- SCHWECHHEIMER C., Understanding gibberellic acid signalling – are we there yet ?, Current Opinion in Plant Biology, 11: 9-15, 2008
- SHELDON C.C., HILLS M.J., LISTER C., DEAN C., DENNIS E.S., PEACOCK W.J., Resetting of FLOWERING LOCUS C expression after epigenetic repression by vernalization, PNAS, 105:2214-2219, 2008
- SMALWOOD M., BOWLES D.J., Plants in a cold climate, The Philosophical Transactions of the Royal Society, 357 :831-847, 2002
- SOBRINO B., BRION M., CARRACEDO A., SNP in forensic genetics: a review on SNP typing methodologies, Forensic Science International, 154:181-194, 2005
- SUNG S., AMASINO R.M., Vernalization in Arabidopsis thaliana is mediated by the PHD finger protein VIN3, Nature, 427:159-164, 2004
- SUNG S., AMASINO R.M., Remembering winter: toward a molecular understanding of vernalization, Annual Review of Plant Biology, 56:491-508, 2005
- SUNG S., AMASINO R.M., Molecular genetic studies of the memory of winter, Journal of Experimental Botany, 57:3369-3377, 2006
- THOMAS B., Light signals and flowering, Journal of Experimental Botany, 57:3387-3393, 2006
- TILL B.J., ZERR T., COMAI L., HENIKOFF S., A protocol for TILLING and Ecotilling in plants and animals, Nature Protocols, 1:2465-2477, 2006
- VISWANATHAN C., ZHU J-K, Molecular genetic analysis of cold-regulated gene transcription, The Philosophical Transactions of the Royal Society, 357:877 :886, 2002
- WERLE E., SCHNEIDER C., RENNER M., VOLKER M., FIEHN W., Convenient single-step, one tube purification of PCR products for direct sequencing, Nucleic Acids Research, 22:4354-4355, 1994
- ZHU C., GORE M., BUCKLER E.S., YU J., Status and prospects of association mapping in plants, The Plant Genome, 1:5-20, 2008

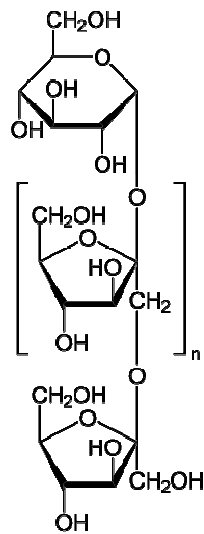


Figure 1 : Structure de l'inuline (Cornelis, 2003)

Ce polymère de β -2,1-fructose est stocké dans les racines de chicorée en tant que réserve d'énergie.



Figure 2 : *Cichorium intybus* L. (Cornelis, 2003)

La chicorée industrielle peut atteindre 80-90cm de haut. Elle présente des feuilles oblongues et lancéolées formant une rosette organisée en spirale au niveau de la tige. Lors de sa phase reproductive, elle forme de petites fleurs bleues.

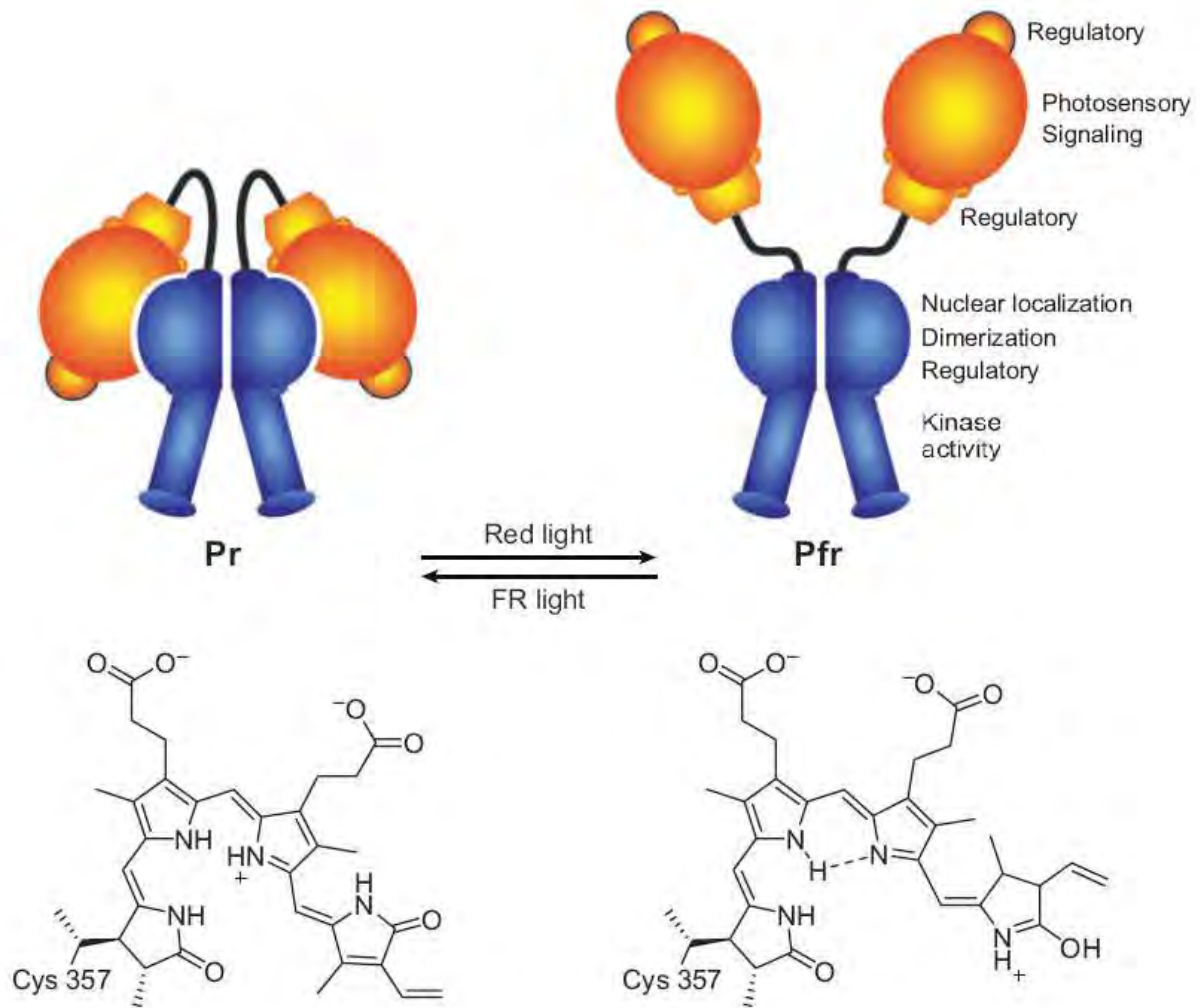


Figure 3 : Structure et fonction des phytochromes (Bae et al., 2008)

Les phytochromes sont des protéines dimériques se présentant selon deux conformations différentes. Le passage de la forme Pr inactive à la forme Pfr active est induit par l'absorption de lumière rouge au niveau de la phytochromobiline. Cette photo-isomérisation entraîne des modifications structurales se traduisant par l'exposition des domaines N-terminal et C-terminal permettant aux phytochromes d'interagir avec les autres protéines impliquées dans la transduction du signal.

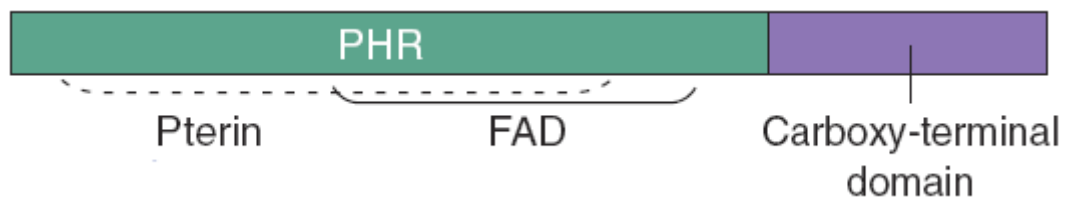


Figure 4 : Structure des cryptochromes (Lin, 2005)

Les cryptochromes sont des protéines très similaires aux photolyases. Ils possèdent en effet une extrémité N-terminale présentant un domaine PHR (photolyase-related) pouvant lier des chromophores, tels qu'une FAD et une ptérine. L'extrémité C-terminale des cryptochromes est quant à elle différente de celle des photolyases.

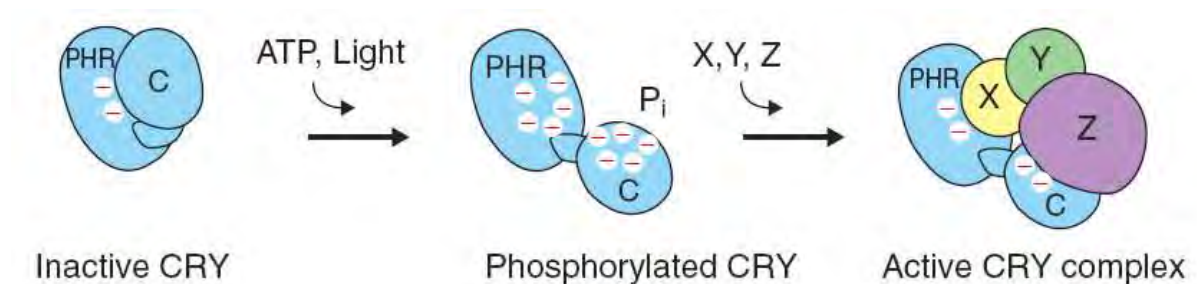


Figure 5 : Mécanisme d'action des cryptochromes (Lin, 2005)

L'extrémité C-terminale des cryptochromes est phosphorylée en réponse à la lumière bleue par l'intermédiaire d'un transfert de phosphate provenant de l'ATP lié au niveau du domaine PHR de l'extrémité N-terminale. Étant donné que ce domaine amino-terminal est chargé négativement, le domaine carboxy-terminal phosphorylé va être repoussé, entraînant ainsi un changement de conformation permettant aux cryptochromes d'interagir avec d'autres protéines impliquées dans la transduction du signal.

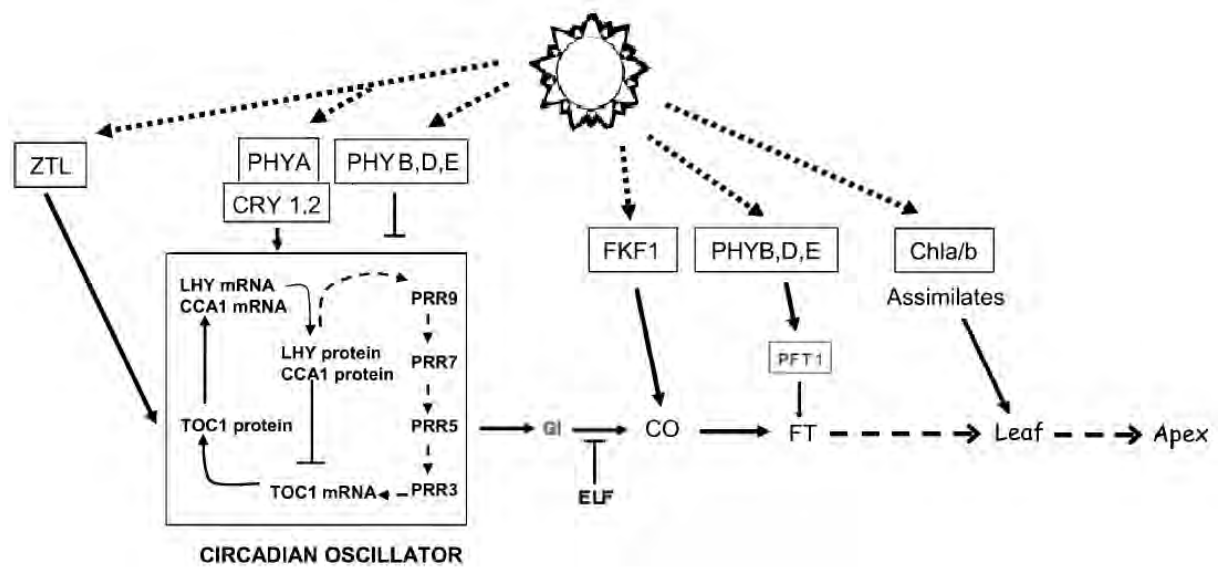


Figure 6 : Voie de floraison dépendante de la photopériode (Thomas, 2006)

Les photorécepteurs ont pour fonction de synchroniser l'horloge circadienne aux cycles quotidiens du jour et de la nuit. Cette boucle de feedback négatif régule ensuite l'expression des gènes GI et CO dont la fonction est d'activer la transcription du gène FT. La protéine FT va finalement initier le développement floral via l'activation des gènes d'identité florale au niveau du méristème apical.

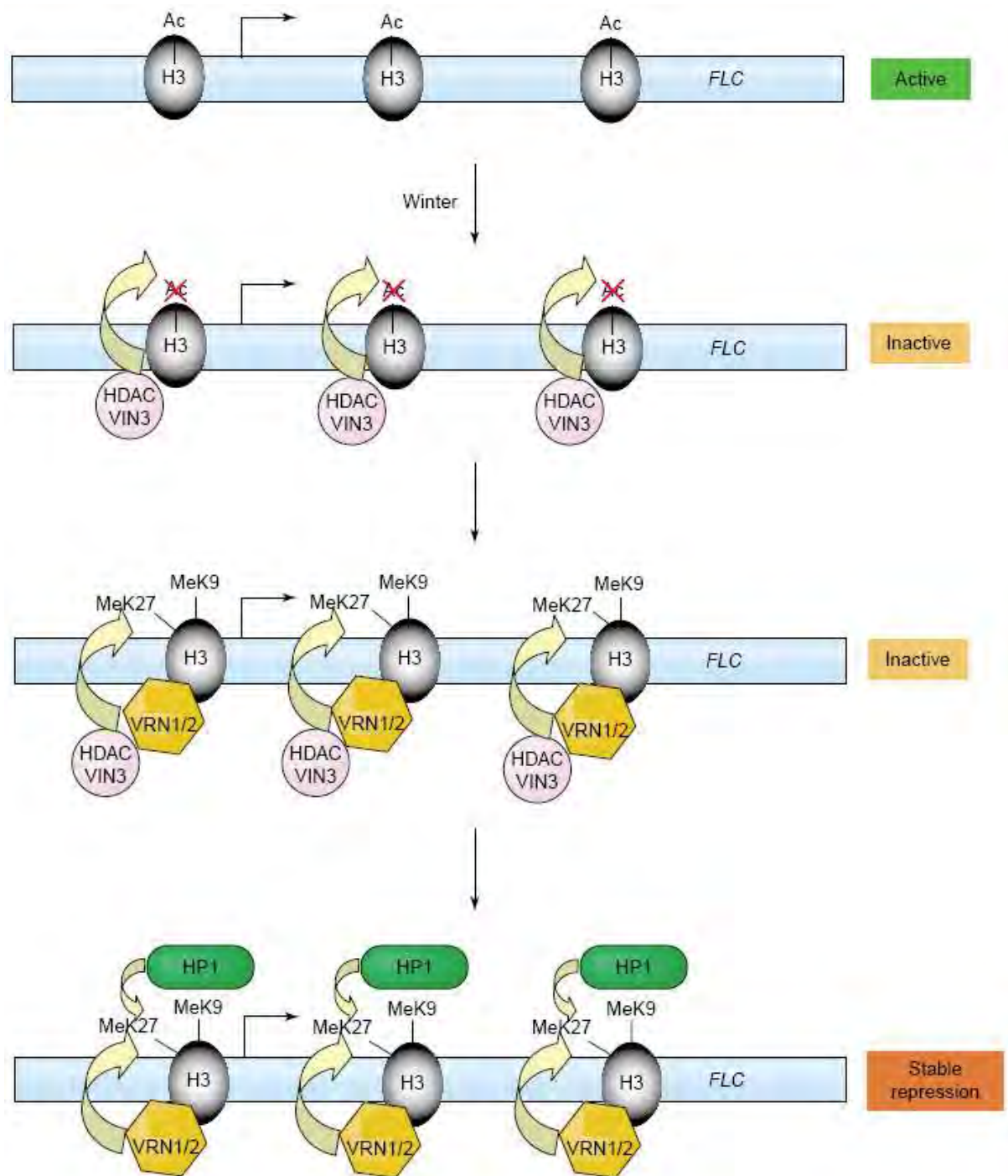


Figure 7 : Régulation épigénétique du locus *FLC* suite à la vernalisation (Sung et al., 2004)

*La vernalisation induit une variation du « code des histones » au niveau du locus *FLC*. L'exposition au froid induit l'expression du gène *VIN3*. La protéine *VIN3* semble nécessaire à l'action du complexe de déacétylation des histones (*HDAC*) qui agit au niveau de l'histone 3 de la chromatine de *FLC*. Cette déacétylation induit le recrutement des protéines *VRN1* et *VRN2* qui sont impliquées dans la méthylation au niveau de l'histone 3 de la chromatine de *FLC*. Toutes ces modifications induisent la conversion de l'euchromatine du gène *FLC* en hétérochromatine, inhibant ainsi son expression. La répression de *FLC* peut être maintenue, au cours des mitoses, par la liaison de protéines telles qu'*HP1* reconnaissant spécifiquement les modifications de la chromatine, la rendant ainsi inaccessible à la machinerie de transcription.*

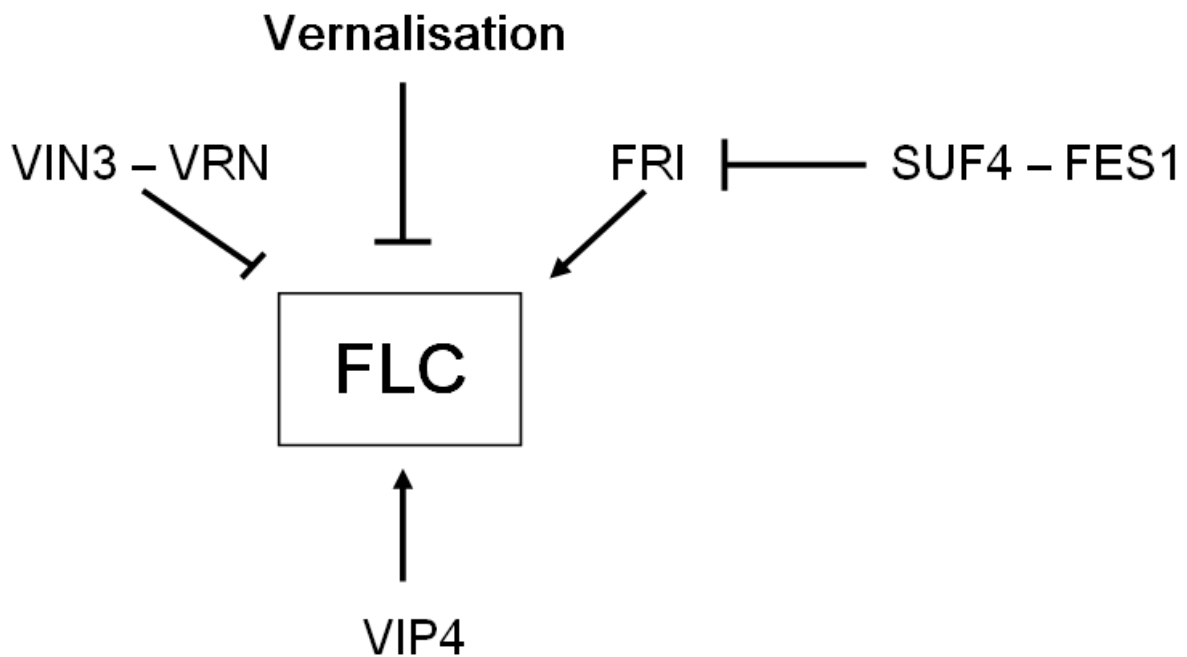


Figure 8 : Voie de floraison dépendante de la vernalisation (Boss et al., 2004)

*Le gène **FLC** est régulé négativement par le froid. Cette répression est liée à la régulation épigénétique du gène **FLC** par les protéines **VIN3** et **VRN**. Les gènes **FRI** et **VIP** régulent positivement l'expression du gène **FLC**.*

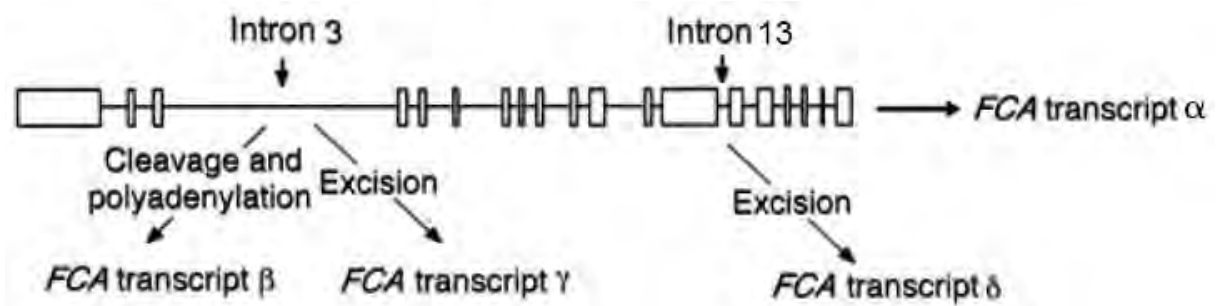


Figure 9 : Epissage alternatif de l'ARNm du gène *FCA* (Macknight et al., 2002)

*L'ARNm du gène **FCA** peut être traité à deux positions différentes, engendrant alors quatre transcrits différents. La présence de l'intron 3 correspond au transcrit α. Un clivage prématuré et une polyadénylation de l'intron 3 donne le transcrit β, tandis que l'excision totale de l'intron 3 donne le transcrit γ qui est le seul à encoder la protéine **FCA** active dans le contrôle de la floraison. L'excision de l'intron 13 produit quant à elle le transcrit δ.*

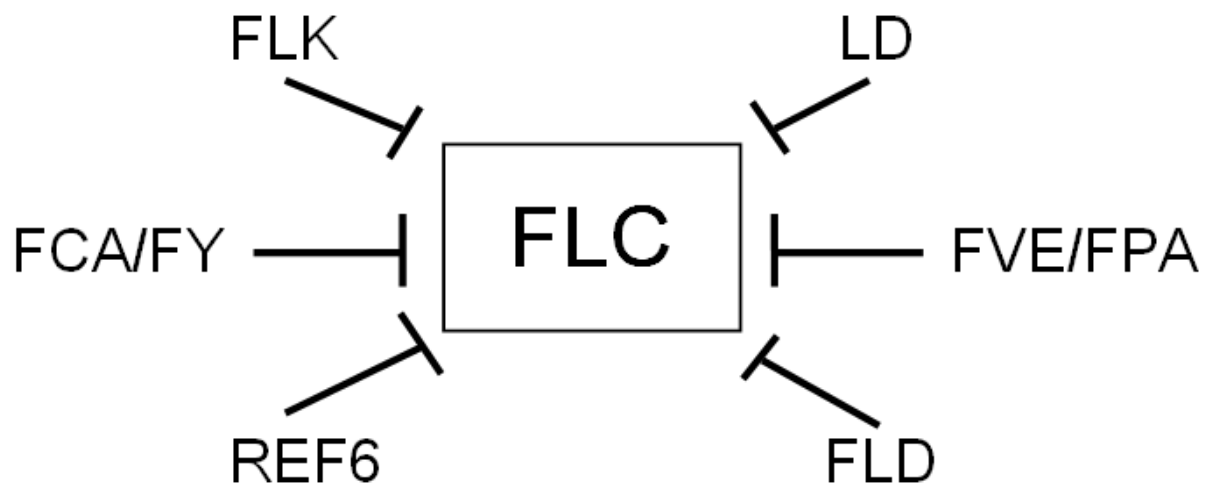


Figure 10 : Voie de floraison autonome (Boss et al., 2004)

Les protéines de liaison à l'ARN telles que FCA, FY, FPA, FLK et LD, et les protéines impliquées dans le remodelage de la chromatine telles que FLD, FVE et REF6 ont pour fonction de réguler négativement l'expression du gène FLC, agissant ainsi en parallèle de la vernalisation.

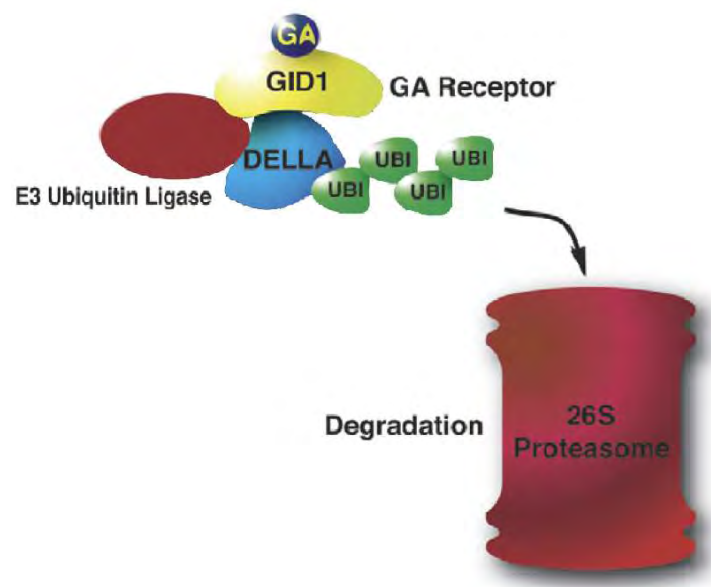


Figure 11 : Régulation des protéines DELLA par l'acide gibbérellique (Schwechheimer, 2008)

Lorsque l'acide gibbérellique se lie à son récepteur GID1, il induit l'interaction de celui-ci avec les protéines DELLA. Ce processus déclenche l'ubiquitination des protéines DELLA par une E3 ubiquitine ligase, induisant ainsi leur dégradation par le protéasome.

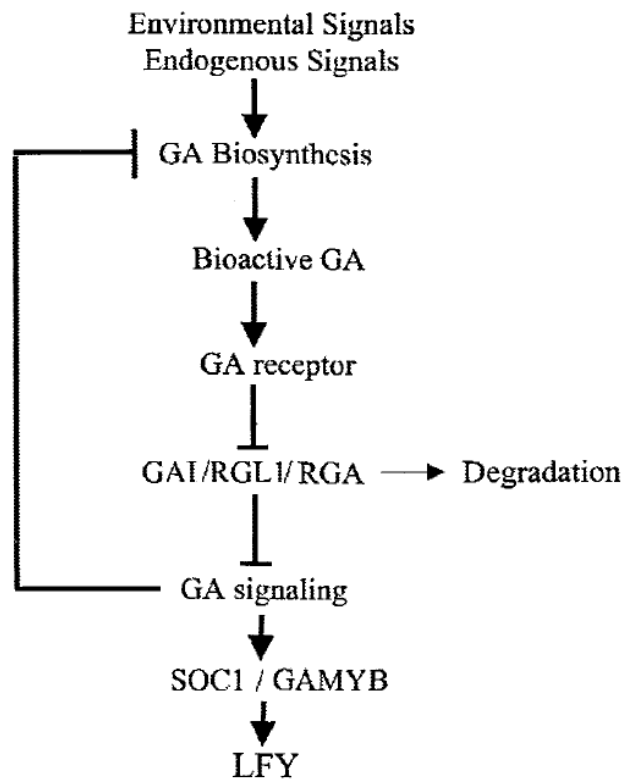


Figure 12 : Voie de floraison dépendante de l'acide gibbérellique (Mouradov et al., 2002)

La biosynthèse de l'acide gibbérellique est induite tant par des signaux endogènes qu'exogènes. Une fois lié à son récepteur, il induit la dégradation des protéines DELLA inhibant la voie de signalisation. L'acide gibbérellique peut alors réguler l'expression des gènes en aval, tels que SOC1 et LFY, ainsi que sa propre expression.

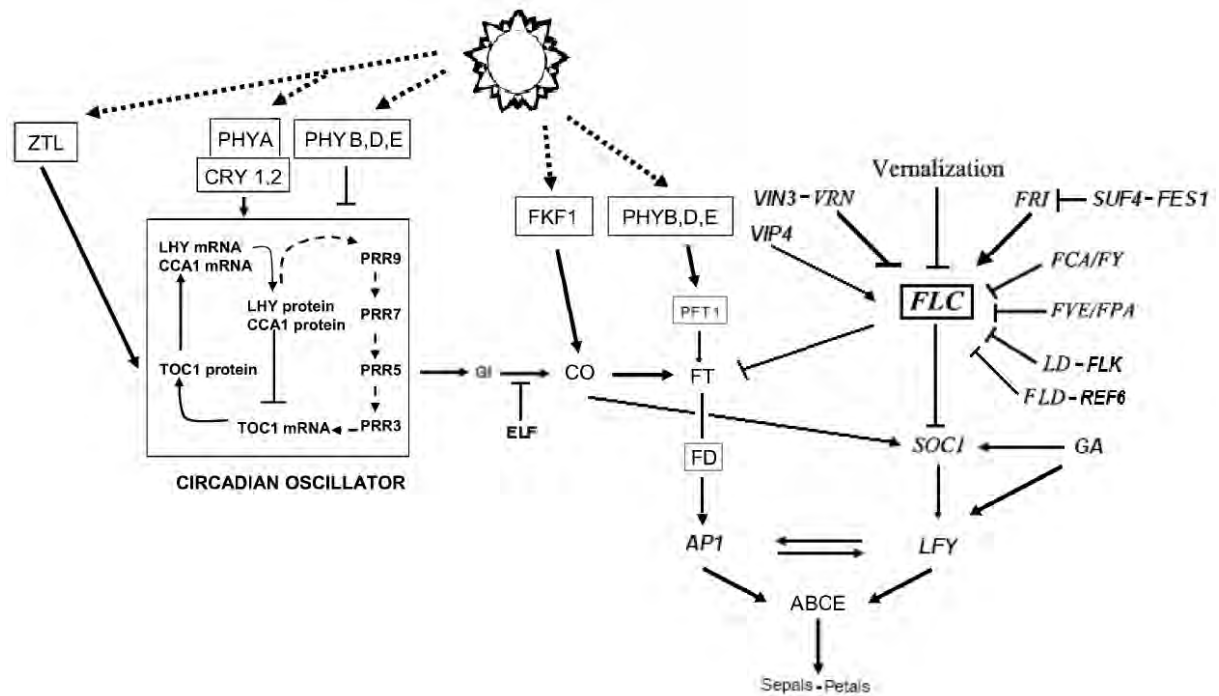


Figure 13 : Intégration des différentes voies de floraison

Les différentes voies de floraison convergent vers la régulation des mêmes gènes en aval. Le gène *FT* intègre ainsi les signaux de la voie dépendante de la photopériode, alors que le gène *FLC* intègre les signaux de la voie dépendante de la vernalisation et autonome. Le gène *SOC1* est régulé positivement par le gène *CO* et la voie dépendante de l'acide gibbérélique, et régulé négativement par le gène *FLC*. Le gène *LFY* est régulé positivement par le gène *SOC1* et la voie dépendante de l'acide gibbérélique.

Les gènes d'identité du méristème floral (*LFY* – *AP1*) encodent des facteurs de transcription dont la fonction est de réguler l'expression des gènes d'identité des organes floraux (gènes *ABCE*) qui permettent finalement de spécifier les différentes parties de la fleur.

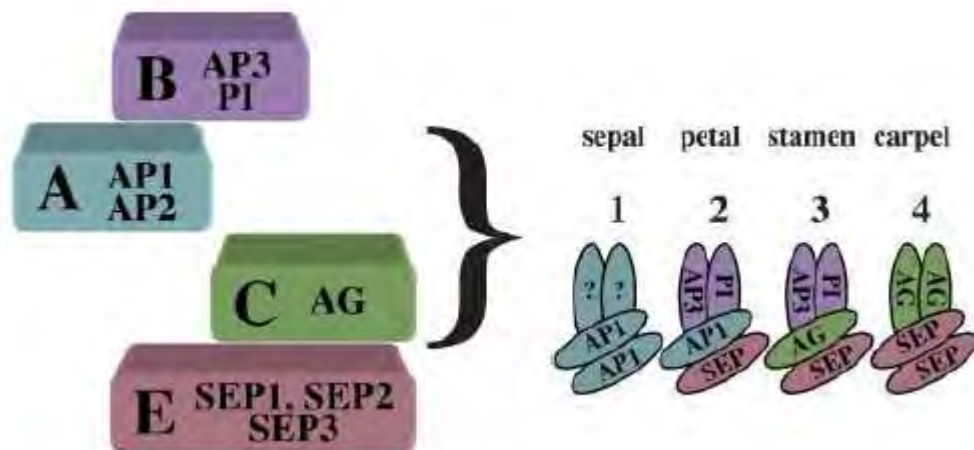


Figure 14 : Modèle ABCE à la base du développement floral (Jack et al., 2004)

L'interaction des différentes classes de gènes d'identité des organes floraux induit le développement des divers organes composant la fleur

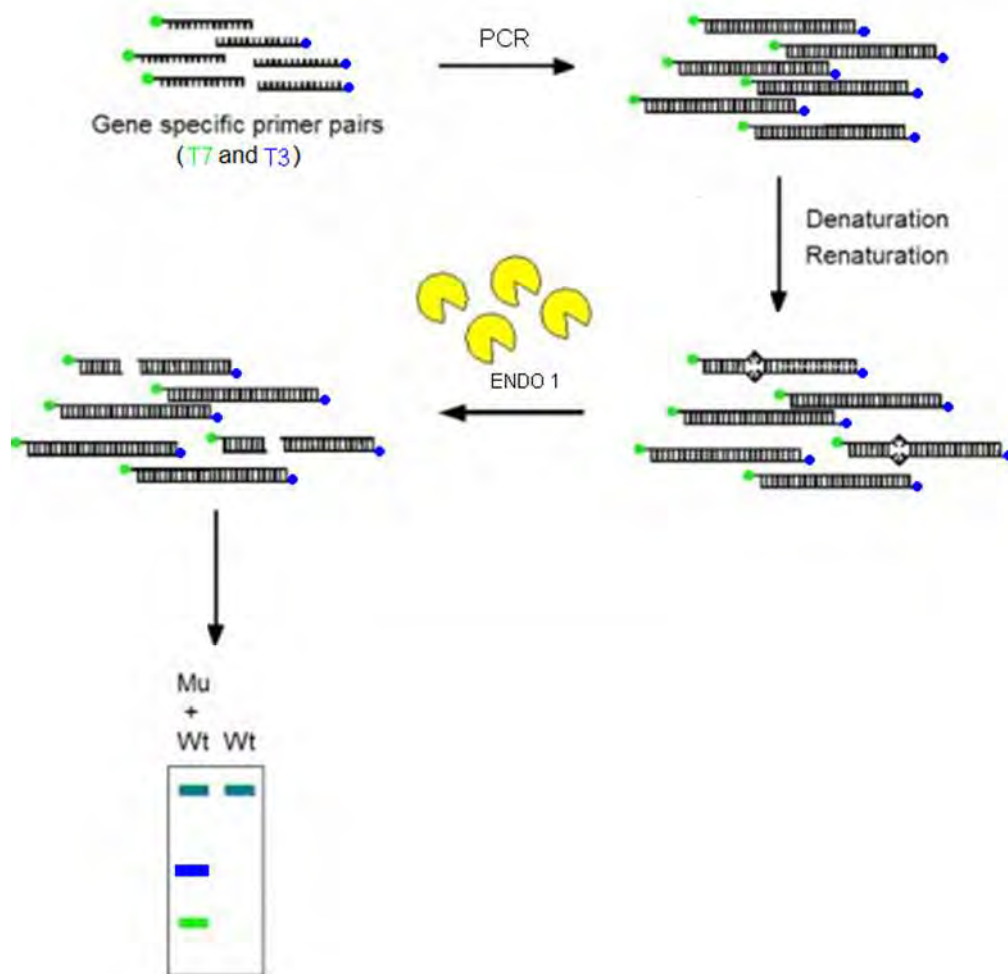


Figure 15 : Principe de la technique d'EcoTILLING (http://www.risoe.dk/rispubl/PRD/AR2003/plr_Eco-Tilling.htm)

Une PCR à l'aide d'amorces spécifiques de gènes marquées différemment est réalisée. Après amplification, les échantillons sont dénaturés puis réhybridés, induisant ainsi la formation d'hétéroduplexes entre les brins d'ADN polymorphes. Les mésappariements dus aux SNP sont ensuite partiellement clivés par une endonucléase spécifique de simple brin (Endo-I). Les fragments fluorescents résultant de cette digestion sont finalement séparés par électrophorèse, ce qui permet de localiser et de dénombrer les SNP.

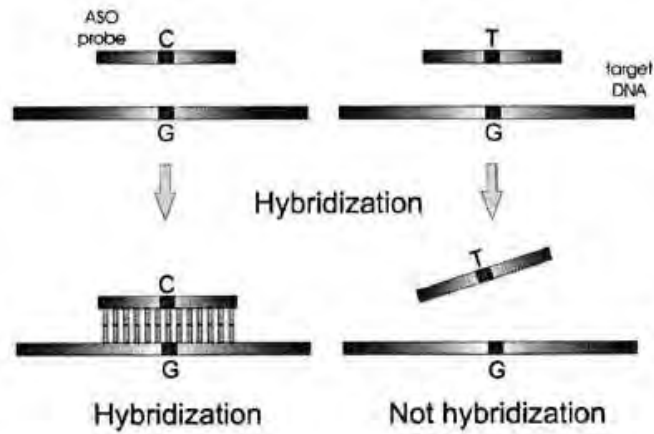


Figure 16 : Hybridation d'oligonucléotide (Sobrino et al., 2004)

Le génotypage du SNP est basé sur l'hybridation stable de la sonde complémentaire spécifique de l'allèle

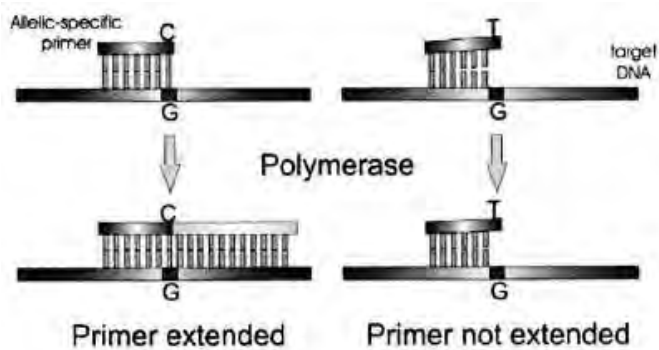


Figure 17 : Extension d'amorce (Sobrino et al., 2004)

Le génotypage du SNP est basé sur la capacité de la polymérase à prolonger l'amorce complémentaire de l'allèle et liée de façon stable à la séquence

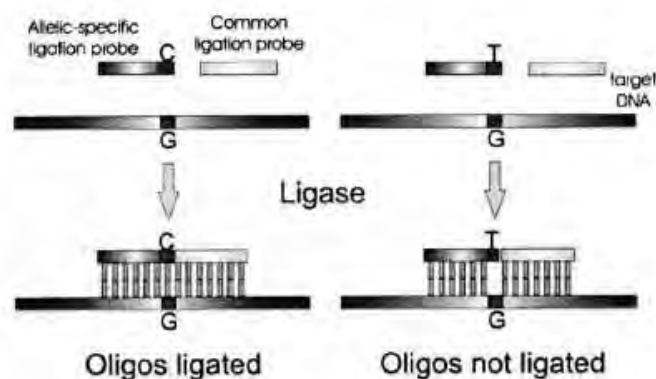


Figure 18 : Ligation d'oligonucléotide (Sobrino et al., 2004)

Le génotypage du SNP est basé sur la capacité de la ligase à assembler deux sondes adjacentes complémentaires de l'allèle et liées de façon stable à la séquence

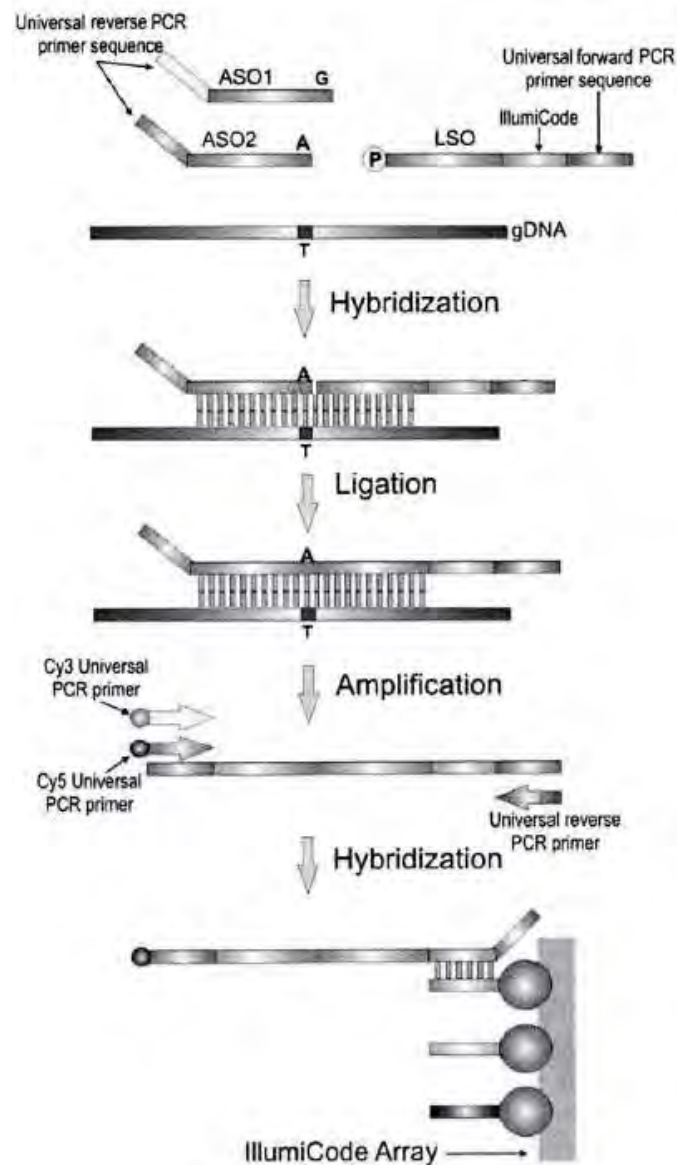


Figure 19 : Variante du système de Golden Gate Assay (Sobrino et al., 2004)

L'ASO complémentaire va s'hybrider de manière stable à la séquence avant d'être reliée à la LSO adjacente par la ligase. Le produit de ligation est ensuite amplifié à l'aide des amorces universelles avant d'être hybridé sur un support solide à l'aide des séquences ZipCodes. Le génotypage du SNP est basé sur la détection du fluorochrome associé à l'amorce universelle (RV) spécifique de l'allèle.

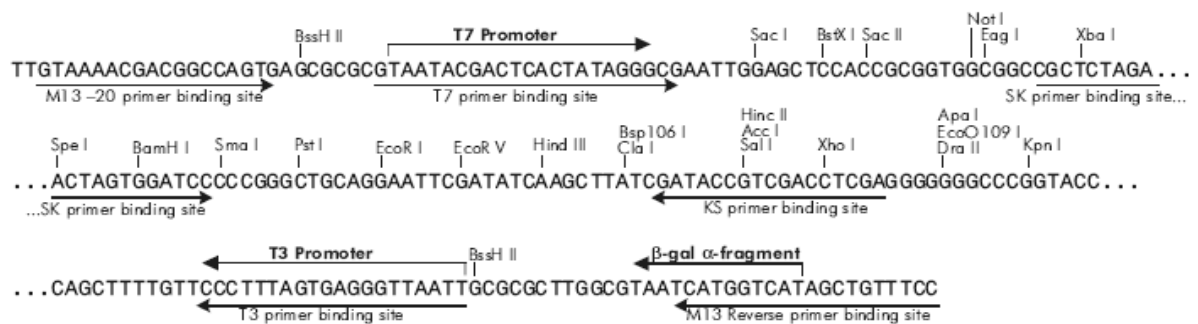


Figure 20 : Séquence du plasmide pBluescript II KS+

Les séquences universelles T3 et T7 sont localisées de part et d'autre du site multiple de clonage (MCS), à une cinquantaine de paires de bases des sites de restriction XhoI et XbaI.

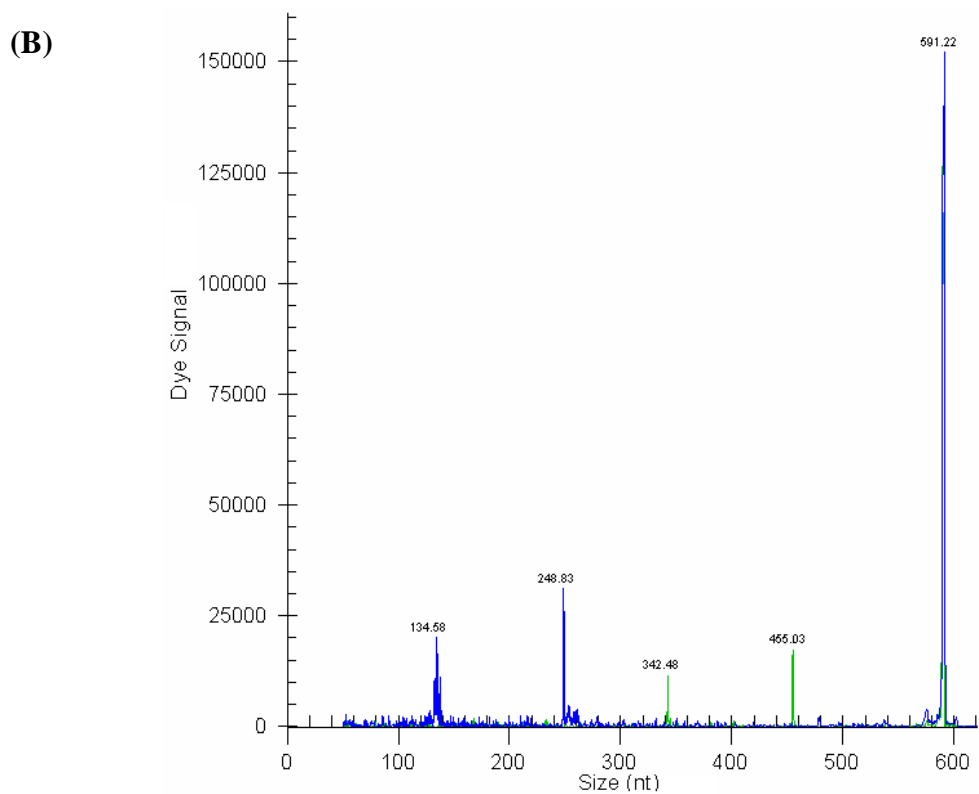
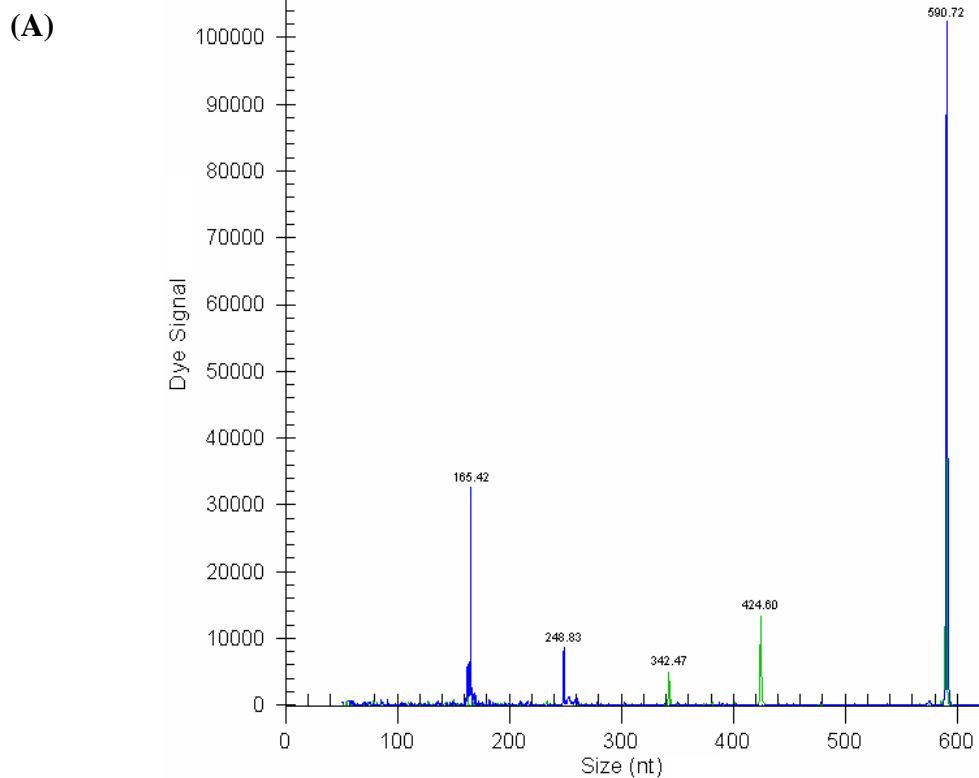


Figure 21 : Sortie de l'analyse EcoTILLING pour les clones bactériens 17 (A) et 22 (B) présentant la séquence de chicorée homologue au gène candidat ELF4

Une complémentarité entre les pics bleus et verts détectés dans chaque séquence est clairement apparue lors de cette analyse.

Analyses de séquençage

La légende ci-dessous explique les différentes données reprises dans l'ensemble des tableaux et figures illustrant la section 6 des résultats (Analyses de séquençage)

Figure (A) : Localisation des SNP dans la séquence ADN consensus de chicorée ¹

Figure (B) : Localisation de domaines conservés dans la séquence ADN consensus de chicorée ²

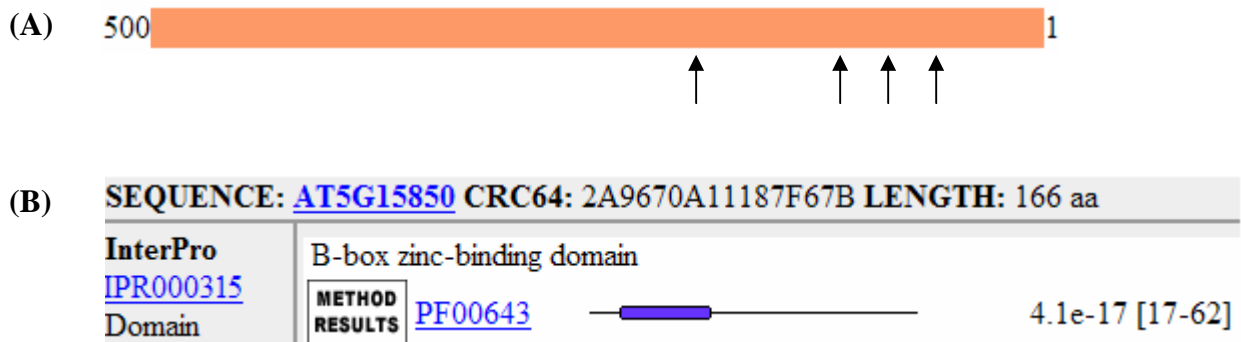


Figure 22 : Résultats de séquençage pour le gène candidat AT5G15850 (COL)

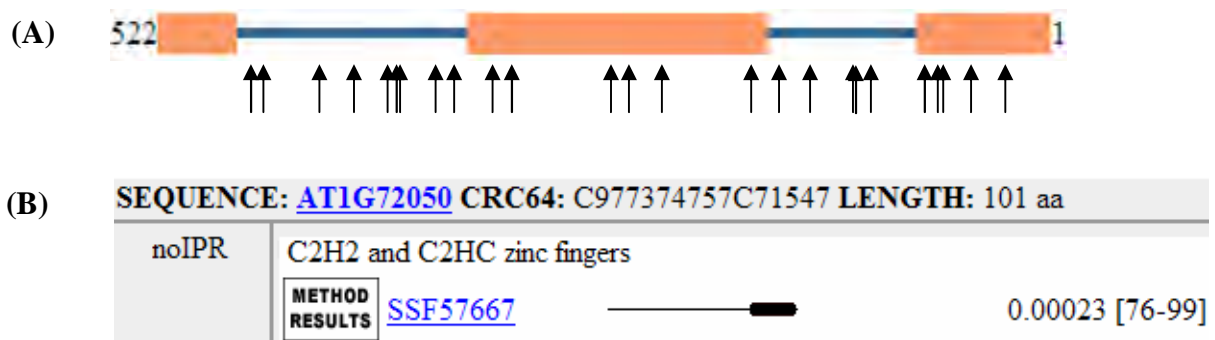


Figure 23 : Résultats de séquençage pour le gène candidat AT1G72050 (ELF6)

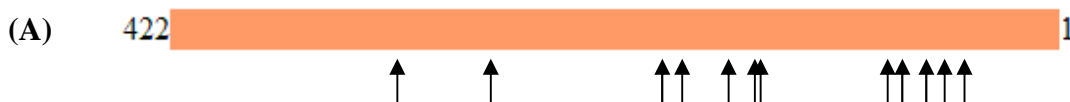


Figure 24 : Résultats de séquençage pour le gène candidat AT3G24440 (VIL1)

¹ <http://www.ncbi.nlm.nih.gov/spidey/>

² <http://www.bioinformatics.nl/iprscan/iprscan>

	270	270	280	290	300	310	320	330	340	350	360
036-03-contig	150	TCGCAATGTGGTCCAATTGACTGATTATTTCTC	TTTGTGTGATTGATCTACGCAGATCATCAATGATCGCGAGACAGGCAGATCCAGGGGGTTTGGTT								
036-22-contig	270	TCGCAATGTGGTCCAATTGACTGATTATTTCTC	TTTGTGTGATTGATCTACGCAGATCATCAATGATCGCGAGACAGGCAGATCCAGGGGGTTTGGTT								
036-52-contig	270	TCGCAATGTGGTCCAATTGACTGATTATTTCTC	TTTGTGTGATTGATCTACGCAGATCATCAATGATCGCGAGACAGGCAGATCCAGGGGGTTTGGTT								
	230	230	240	250	260	270	280	290	300	310	320
036-05-contig	230	CCTCTTCTCTTATAAGAGACGACGATCGATCGATTTTATGATGCAGTTTCAATTTATTG	ATTACAAATTGTTGATTAGGATCCTGATGTA	TTTTTGTG							
036-32-contig	230	CCTCTTCTCTTATAAGAGACGACGATCGATCGATTTTATGATGCAGTTTCAATTTATTG	ATTACAAATTGTTGATTAGGATCCTGATGTA	TTTTTGTG							
036-28-contig	230	CCTCTTCTCTTATAAGAGACGACGATCGATCGATTTTATGATGCAGTTTCAATTTATTG	ATTACAAATTGTTGATTAGGATCCTGATGTA	TTTTTGTG							
036-46-contig	90	CCTCTTCTCTTATAAGAGACGACGATCGATCGATTTTATGATGCAGTTTCAATTTATTG	ATTACAAATTGTTGATTAGGATCCTGATGTA	TTTTTGTG							
	160	160	170	180	190	200	210	220	230	240	250
036-03-contig	25	T-TTACTATGGTTT-CTTGATCTGTTACTGTTTCTGTTACTTTG	---GATATCTATC	ACGGTACGTTT	CATCCTCTTT	TTTCTTAT	--GAGGTGGCGATCGAT				
036-22-contig	145	T-TTACTATGGTTT-CTTGATCTGTTACTGTTTCTGTTACTTTT	---GATATCTATC	ACGGTACGTTT	CATCCTCTTT	TTTCAAT	--GAGGTGGCGATCGAT				
036-05-contig	160	TGTTACTATGGTTTCTTGATCTGTTACTATTTGTGTTACTTTT	ACTGATATCTAAA	ACGGTACGTTT	CATCCTCTTCT	CTTTATA	AGAGACGACGATCGAT				
036-32-contig	160	TGTTACTATGGTTTCTTGATCTGTTACTATTTGTGTTACTTTT	ACTGATATCTAAA	ACGGTACGTTT	CATCCTCTTCT	CTTTATA	AGAGACGACGATCGAT				
036-28-contig	160	TGTTACTATGGTTTCTTGATCTGTTACTATTTGTGTTACTTTT	ACTGATATCTAAA	ACGGTACGTTT	CATCCTCTTCT	CTTTATA	AGAGACGACGATCGAT				
036-46-contig	20	TGTTACTATGGTTTCTTGATCTGTTACTATTTGTGTTACTTTT	ACTGATATCTAAA	ACGGTACGTTT	CATCCTCTTCT	CTTTATA	AGAGACGACGATCGAT				
036-52-contig	145	T-TTACTATGGTTT-CTTGATCTGTTACTGTTTCTGTTACTTTT	---GATATCTATC	ACGGTACGTTT	CATCCTCTTT	TTTCAAT	--GAGGTGGCGATCGAT				

Figure 25 : Résultats de séquençage pour le gène candidat AT1G79730.1 (ELF7)

Le premier alignement présente trois séquences différentes et le second alignement présente quatre séquences différentes. Le troisième alignement présente les sept séquences différentes et montre les deux populations de séquences paralogues.

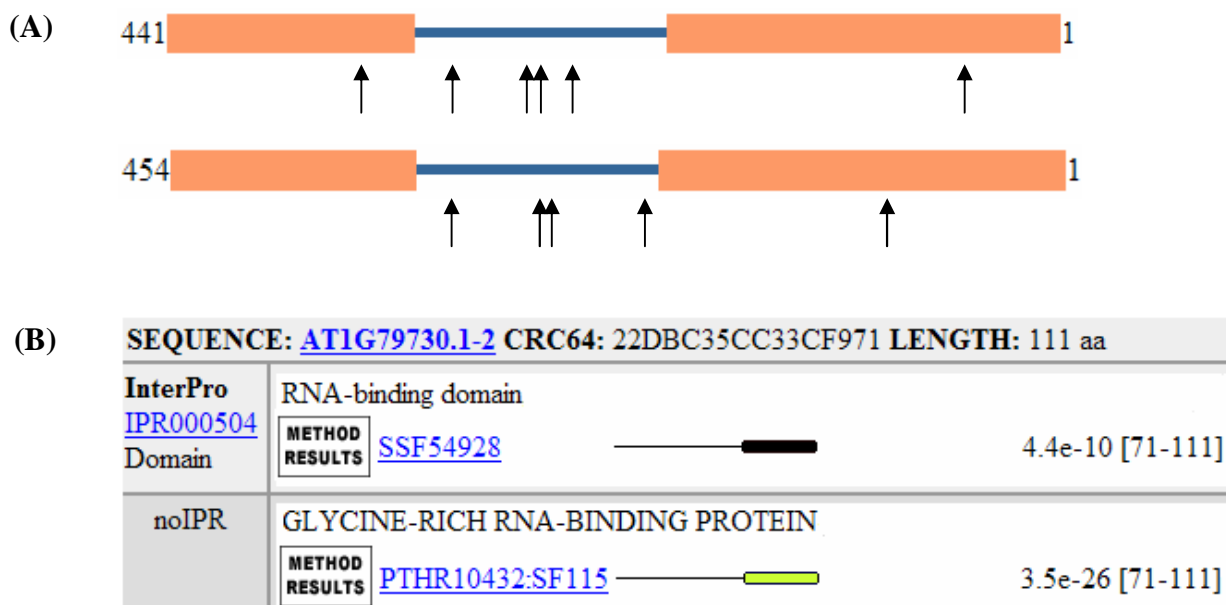


Figure 26 : Résultats de séquençage pour le gène candidat AT1G79730.1 (ELF7)



Figure 27 : Résultats de séquençage pour le gène candidat AT5G57380.1 (VIN3)



Figure 28 : Résultats de séquençage pour le gène candidat AT5G65070 (FLC)

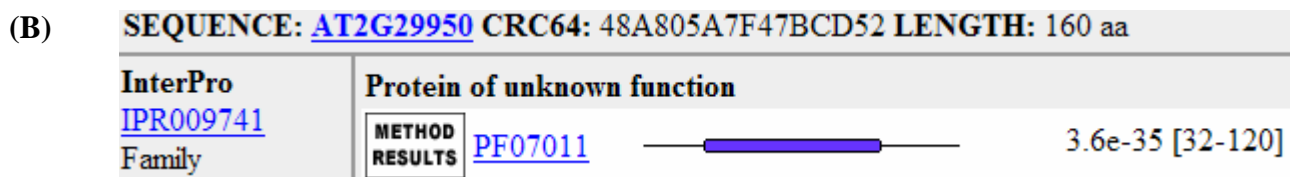


Figure 29 : Résultats de séquençage pour le gène candidat AT2G29950 (ELF4)

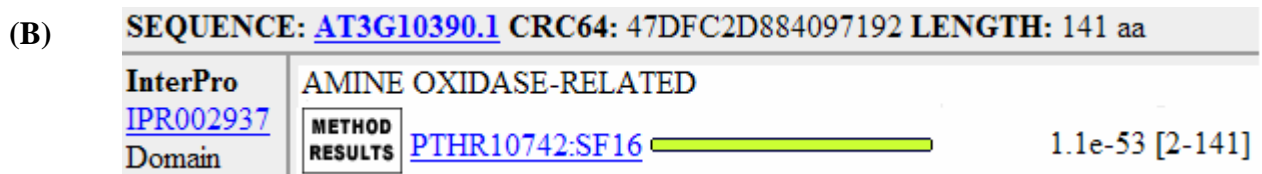


Figure 30 : Résultats de séquençage pour le gène candidat AT3G10390.1 (FLD)

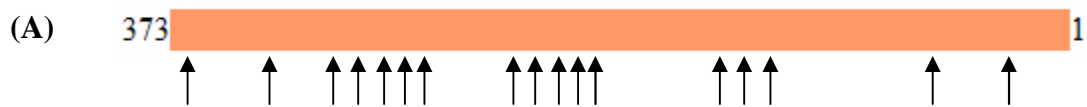


Figure 31 : Résultats de séquençage pour le gène candidat AT1G22770.1 (GI)

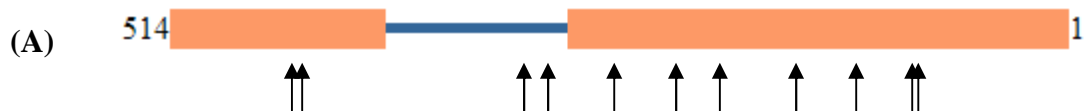


Figure 32 : Résultats de séquençage pour le gène candidat AT5G15840.1 (CO)

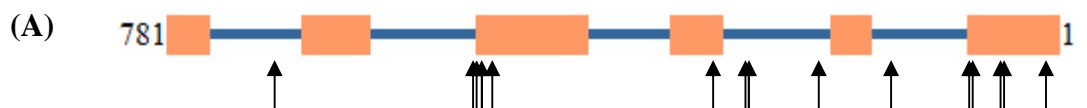


Figure 33 : Résultats de séquençage pour le gène candidat AT5G65070 (FLC)



Figure 34 : Résultats de séquençage pour le gène candidat AT1G04400.1 (CRY2)



Figure 35 : Résultats de séquençage pour le gène candidat AT1G25540.1 (PFT1)

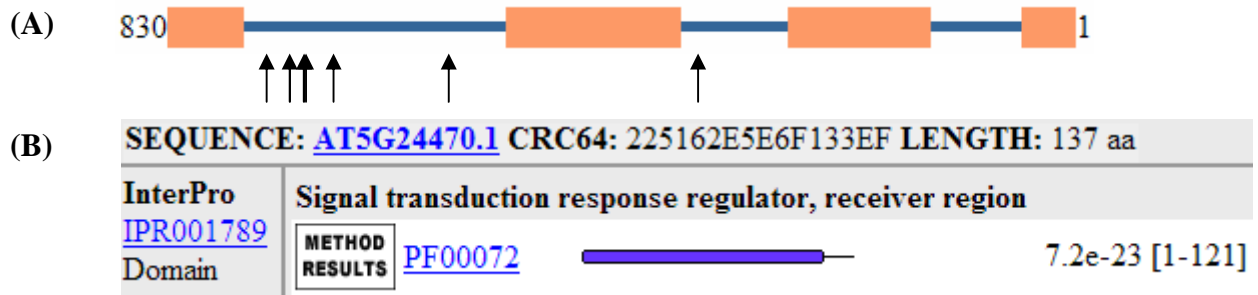


Figure 36 : Résultats de séquençage pour le gène candidat AT5G24470.1 (APRR5)

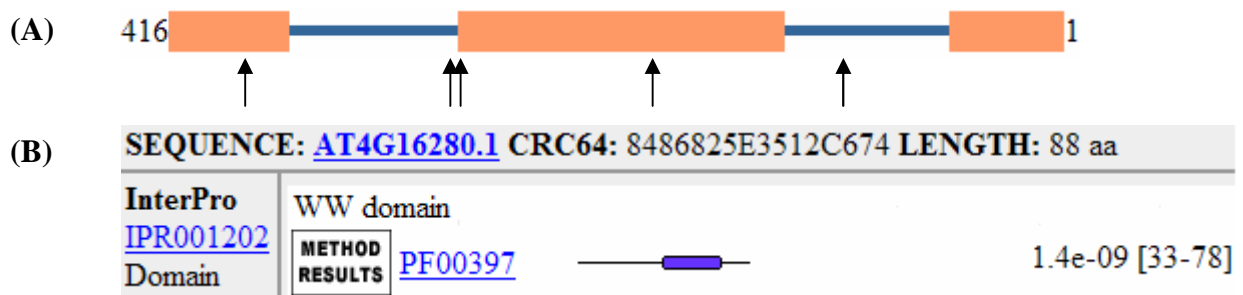


Figure 37 : Résultats de séquençage pour le gène candidat AT4G16280.1 (FCA)